

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 16, 2002, 23:19:15 ; Search time 4188 Seconds  
(without alignments)  
16601.386 Million cell updates/sec

Title: US-08-961-083-55

Perfect score: 2389

Sequence: 1 TTCTTAGCAGGTGGAGCTGT.....TAAGTAAGGAAAAATTAAC 2389

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:\*  
1: gb\_ba:\*  
2: gb\_htg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_pro:\*  
11: gb\_scs:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_pro:\*  
27: em\_sts:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_htg\_hum:\*  
31: em\_htg\_huv:\*  
32: em\_htg\_other:\*  
33: em\_htg\_mus:\*  
34: em\_htg\_pln:\*  
35: em\_htg\_fod:\*  
36: em\_htg\_mam:\*  
37: em\_htg\_vrt:\*  
38: em\_sy:\*  
39: em\_htgo\_hum:\*  
40: em\_htgo\_mus:\*  
41: em\_htgo\_other:\*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2388	100.0	2389	6	ARI20265	ARI20265 Sequence
2	2388	100.0	2341	1	AF291695	AF291695 Streptococ
3	2388	100.0	8195	6	BD003774	BD003774 Polynucle
4	2388	100.0	10256	1	AE007418	AE007418 Streptococ
5	2365.6	99.0	10320	1	AE008479	AE008479 Streptococ
6	2328.2	97.5	20035	2	SPNEU1915	SPNEU1915 Streptococ
7	1434.4	60.0	2535	1	AF340221	AF340221 Streptococ
8	1374.2	57.5	2523	6	AX343072	AX343072 Sequence
9	1374.2	57.5	2647	6	AX343073	AX343073 Sequence
10	991	41.5	2457	1	AF318954	AF318954 Streptococ
11	990	41.4	2517	1	AF318955	AF318955 Streptococ
12	990	41.4	11931	1	AE007403	AE007403 Streptococ
13	988.4	41.4	232807	2	SPNEU1901	SPNEU1901 Streptococ
14	987.6	41.3	2290	6	ARI20270	ARI20270 Sequence
15	980.8	41.1	2639	6	AX343074	AX343074 Sequence
16	962.8	40.3	12372	1	AE008464	AE008464 Streptococ
17	869.4	36.4	2166	12	AF340222	AF340222 Synthetic
18	835.8	35.0	75874	2	SPNEU1907	SPNEU1907 Streptococ
19	674	28.2	702	2	SPNEU1929	SPNEU1929 Streptococ
20	653.8	27.4	2359	6	BD003923	BD003923 Polynucle
21	385.4	16.1	1342	6	ARI20328	ARI20328 Sequence
22	385.4	16.1	3120	1	AF318956	AF318956 Streptococ
23	385.4	16.1	3120	6	AX343070	AX343070 Sequence
24	385.4	16.1	5048	6	AX343071	AX343071 Sequence
25	385.4	16.1	6867	6	BD003872	BD003872 Polynucle
26	381.4	16.0	973	6	BD004035	BD004035 Polynucle
27	347	14.5	2528	6	AX343078	AX343078 Sequence
28	253.2	10.6	10312	1	AE007404	AE007404 Streptococ
29	243.4	10.2	1684	6	BD003938	BD003938 Polynucle
30	214.6	9.0	2478	6	AX088372	AX088372 Sequence
31	214.6	9.0	10825	1	AE006623	AE006623 Streptococ
32	213	8.9	2478	1	AB073859	AB073859 Streptococ
33	213	8.9	10618	1	AE010110	AE010110 Streptococ
34	211.4	8.8	1910	1	SAG290952	AJ290952 Streptococ
35	211.4	8.8	2469	6	AX088376	AX088376 Sequence
36	211.4	8.8	3501	1	AF062533	AF062533 Streptococ
37	205	8.6	50354	1	AE014169	AE014169 Streptococ
38	188.2	7.9	492	12	AF340223	AF340223 Synthetic
39	109	4.6	841	6	AR089275	AR089275 Sequence
40	109	4.6	841	6	AR093475	AR093475 Sequence
41	70.6	3.0	115758	9	AC104634	AC104634 Homo sapi
42	63.4	2.7	170522	9	AP002387	AP002387 Homo sapi
43	60.2	2.5	51939	9	AC099341	AC099341 Homo sapi
44	60	2.5	61052	2	AC117074	AC117074 Dictyoste
45	60	2.5	61052	2	AC123513	AC123513 Dictyoste

#### ALIGNMENTS

RESULT 1  
LOCUS ARI20265 2389 bp DNA linear PAT 16-MAY-2001  
DEFINITION Sequence 55 from patent US 6159469.  
ACCESSION ARI20265  
VERSION ARI20265.1 GI:14103841  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 2389)  
AUTHORS Choi,G.H., Kunsch,C.A., Barash,S.C., Dillon,P.J., Dougherty,B.,  
Fannon,M.R. and Rosen,C.A.  
TITLE Streptococcus pneumoniae antigens and vaccines  
JOURNAL Patent: US 6159469-A 55 12-DEC-2000;

```
FEATURES                               Location/Qualifiers
source                                1..2389
                                         /organism="unknown"
BASE COUNT      830 a      461 c      486 g      611 t      1 others
ORIGIN
Query Match      100.0%; Score 2388; DB 6; Length 2389;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2389; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 TTTCTACGCGTTGGGAGCTATCAAGCTGAAAGCGGTTTAAAGAAATTAATCGTTTCCCA 60
Db      1 TTTCTACGCGTTGGGAGCTATCAAGCTGAAAGCGGTTTAAAGAAATTAATCGTTTCCCA 60
QY      61 TATAGATGAAAAACAAGCAGCAAAAAAGAGATTGTACTCTGATGAGGTTAGCAA 120
Db      61 TATAGATGAAAAACAAGCAGCAAAAAAGAGATTGTACTCTGATGAGGTTAGCAA 120
QY      121 GCGTGAAGAAATCAATGCTGAGCAAAATGCTCATCAAGATTAACAGACCAAGGCTATGTAC 180
Db      121 GCGTGAAGAAATCAATGCTGAGCAAAATGCTCATCAAGATTAACAGACCAAGGCTATGTAC 180
QY      181 TTCACATGGCGACCACTATCATTTATACAAATGTAAGGTTCCCTTATGAGCGTATCATCAG 240
Db      181 TTCACATGGCGACCACTATCATTTATACAAATGTAAGGTTCCCTTATGAGCGTATCATCAG 240
QY      241 TGAAGATTACTCATGAAAGATCCAAACTATAGCTAAAGATGAGGATTTGTTATGA 300
Db      241 TGAAGATTACTCATGAAAGATCCAAACTATAGCTAAAGATGAGGATTTGTTATGA 300
QY      301 GGTCAAGGCTGATATGTTTATCAAGGTAGATGAAATTCATGTTTACCTTAAGATGC 360
Db      301 GGTCAAGGCTGATATGTTTATCAAGGTAGATGAAATTCATGTTTACCTTAAGATGC 360
QY      361 TCCCGACGGGATTAAGCTCCGTACAAAAGAGAAATCATCGCAAAAACAAGAGATAG 420
Db      361 TCCCGACGGGATTAAGCTCCGTACAAAAGAGAAATCATCGCAAAAACAAGAGATAG 420
QY      421 TCAACATCGTGAAGGTGAAGTCCAAAGAACATGAGTGTGCTTGCTTGCGACGTTCCGA 480
Db      421 TCAACATCGTGAAGGTGAAGTCCAAAGAACATGAGTGTGCTTGCTTGCGACGTTCCGA 480
QY      481 AGGAGCTTACTACAGATGATGTTATCTTTAATGCTTCTGATATATAGAGATAC 540
Db      481 AGGAGCTTACTACAGATGATGTTATCTTTAATGCTTCTGATATATAGAGATAC 540
QY      541 TGGTATGCTTATATCGTTCCTCATGAGATCATTTACATTCCTTAAGATAGATT 600
Db      541 TGGTATGCTTATATCGTTCCTCATGAGATCATTTACATTCCTTAAGATAGATT 600
QY      601 ATCAGCTAGCGAGTGGCTGCTGTCAGAAAGCCTTCTATCTGTGTCAGAAATCTGTCAAA 660
Db      601 ATCAGCTAGCGAGTGGCTGCTGTCAGAAAGCCTTCTATCTGTGTCAGAAATCTGTCAAA 660
QY      661 TTCAGAAGACCTTATCCGCCGCAAAATAGCGATTAACACTTAAAGAAACAACTGGGTACTTTC 720
Db      661 TTCAGAAGACCTTATCCGCCGCAAAATAGCGATTAACACTTAAAGAAACAACTGGGTACTTTC 720
QY      721 TGTAAAGCAATCCAGAAATCTAATACTTAACACAAGCAACAAGCAAGCACTTAACAGTCA 780
Db      721 TGTAAAGCAATCCAGAAATCTAATACTTAACACAAGCAACAAGCAAGCACTTAACAGTCA 780
QY      781 AGCAAGTCAAAAGTAAATGATGATTAAGTCTCTTGAAGACGCTCTACAAACGCTCTTTGAG 840
Db      781 AGCAAGTCAAAAGTAAATGATGATTAAGTCTCTTGAAGACGCTCTACAAACGCTCTTTGAG 840
QY      841 TCAAGACACATGTAGAAATCTGATGCGCTTCTTTTGTATCCAGACAAATATCAAGTGAAC 900
Db      841 TCAAGACACATGTAGAAATCTGATGCGCTTCTTTTGTATCCAGACAAATATCAAGTGAAC 900
QY      901 AACTGAGAGTGTGAGAGTCCACAGAGAGATCATTTACCACTTCATCCCTTACTCTCAAT 960
Db      901 AACTGAGAGTGTGAGAGTCCACAGAGAGATCATTTACCACTTCATCCCTTACTCTCAAT 960
```

```
QY      961 GTCGTGAATTTGAAGAAGCAATGCTCGTATATATTCCTCCCTTCTGTATCTGTTCAACCATTTG 1020
Db      961 GTCGTGAATTTGAAGAAGCAATGCTCGTATATATTCCTCCCTTCTGTATCTGTTCAACCATTTG 1020
QY      1021 GGTACCAATTTCAAGGCCGGAACCAACCAAGTCCACACCGCATCCGGAACTGATGCCAG 1080
Db      1021 GGTACCAATTTCAAGGCCGGAACCAACCAAGTCCACACCGCATCCGGAACTGATGCCAG 1080
QY      1081 CCCGCAACCTGCAACCAATCTTAAATATGACTCAAAATCTTCTGTTAGTACAGTGGT 1140
Db      1081 CCCGCAACCTGCAACCAATCTTAAATATGACTCAAAATCTTCTGTTAGTACAGTGGT 1140
QY      1141 ACGAAAAGTTGGGGAAGATATGATTTGGAAGAAAAGGCAATCTCTGTTATGCTTTGTC 1200
Db      1141 ACGAAAAGTTGGGGAAGATATGATTTGGAAGAAAAGGCAATCTCTGTTATGCTTTGTC 1200
QY      1201 GAAAGATTTTACCATCTGAAGACTGTTAAATACTTGAAGCAAGTTATCCAAACAAGAGAG 1260
Db      1201 GAAAGATTTTACCATCTGAAGACTGTTAAATACTTGAAGCAAGTTATCCAAACAAGAGAG 1260
QY      1261 TGTTCACACACTTTAAGCTTAAAAAAGAAAATGTTGCTCCGTCGACCAAGAAATTTTA 1320
Db      1261 TGTTCACACACTTTAAGCTTAAAAAAGAAAATGTTGCTCCGTCGACCAAGAAATTTTA 1320
QY      1321 TGATTAAGCATATATCTGTTAAGTGAAGGCTCATAAAGCCTTGTGTAATTAAGGTTGC 1380
Db      1321 TGATTAAGCATATATCTGTTAAGTGAAGGCTCATAAAGCCTTGTGTAATTAAGGTTGC 1380
QY      1381 TAATTCGTATTTCCAAAGCTTATAGCAAAATTAATTAAGACCTTGAATGATCCACTTA 1440
Db      1381 TAATTCGTATTTCCAAAGCTTATAGCAAAATTAATTAAGACCTTGAATGATCCACTTA 1440
QY      1441 TAAAGAAAATTGTATGATGATTTATGSCATTCCTAGACCAATTAACCATCCAGAGCG 1500
Db      1441 TAAAGAAAATTGTATGATGATTTATGSCATTCCTAGACCAATTAACCATCCAGAGCG 1500
QY      1501 ACTTGGCAAAACCAATTCCTCAAAATGAGTATCTGAAGACGAATTCGTATTTGCTCAATT 1560
Db      1501 ACTTGGCAAAACCAATTCCTCAAAATGAGTATCTGAAGACGAATTCGTATTTGCTCAATT 1560
QY      1561 AGCTGATTAAGTATACAGCTCAGATGCTTACATTTTGTATGTAAGCAATGATTAATACGTGA 1620
Db      1561 AGCTGATTAAGTATACAGCTCAGATGCTTACATTTTGTATGTAAGCAATGATTAATACGTGA 1620
QY      1621 TGAAGAGATGATATATGTAACGCTCATATGAGGCAATAGTCACTGATTTGAAAAGATAG 1680
Db      1621 TGAAGAGATGATATATGTAACGCTCATATGAGGCAATAGTCACTGATTTGAAAAGATAG 1680
QY      1681 CCTTTCGTATTAAGAAAAGTTGCAAGCTCAAGCTTATCTTAAAGAAAAGGATATCTTACC 1740
Db      1681 CCTTTCGTATTAAGAAAAGTTGCAAGCTCAAGCTTATCTTAAAGAAAAGGATATCTTACC 1740
QY      1741 TCCATCTCCAGACGCAAGATGTTTAAAGCAATCCAACTGAGATATGTCGACGACTATTTTA 1800
Db      1741 TCCATCTCCAGACGCAAGATGTTTAAAGCAATCCAACTGAGATATGTCGACGACTATTTTA 1800
QY      1801 CAATCGTGTGAAGAGGGGAAAAGCAATTCACCTGTTGACCTTCATATATATGTTGAGCA 1860
Db      1801 CAATCGTGTGAAGAGGGGAAAAGCAATTCACCTGTTGACCTTCATATATATGTTGAGCA 1860
QY      1861 TACAGTTGAGGTTTAAAAAAGGTTAATTTGATTTTCTCTCAATTAAGGATCATTTACCATAATAT 1920
Db      1861 TACAGTTGAGGTTTAAAAAAGGTTAATTTGATTTTCTCTCAATTAAGGATCATTTACCATAATAT 1920
QY      1921 TAAATTTGCTGTTGATGATCACACATACAAAGTCCAAATGGCTATACCTTGGAAAGA 1980
Db      1921 TAAATTTGCTGTTGATGATCACACATACAAAGTCCAAATGGCTATACCTTGGAAAGA 1980
QY      1981 TTTGTTTGGAGATTTAATTAAGTATAGTATAGCAACCTTGACGAAGCTGCACATTCCTAATGA 2040
Db      1981 TTTGTTTGGAGATTTAATTAAGTATAGTATAGCAACCTTGACGAAGCTGCACATTCCTAATGA 2040
```

QY 2041 TGGATGGGGCAATGCCAGTGAACATGTGTAGGCAAGAAAGACACAGTGAATCCAAA 2100  
 |||  
 Db 2041 TGGATGGGGCAATGCCAGTGAACATGTGTAGGCAAGAAAGACACAGTGAATCCAAA 2100  
 QY 2101 TAGAATCTTCAAGCGGATGTAAGGCGCAATGAGAGAAACACCTCTGAGCCAGAATCCC 2160  
 |||  
 Db 2101 TAAAGATCTTCAAGCGGATGTAAGGCGCAATGAGAGAAACACCTCTGAGCCAGAATCCC 2160  
 QY 2161 TCAAGTGAAGCTGAAAGAGTGAAGGCGCAATGAGAGAAACACCTCTGAGCCAGAATCCC 2220  
 |||  
 Db 2161 TCAAGTGAAGCTGAAAGAGTGAAGGCGCAATGAGAGAAACACCTCTGAGCCAGAATCCC 2220  
 QY 2221 AGTAAAGGATCTAGTGTGAAGGCGCAATGAGAGAAACACCTCTGAGCCAGAATCCC 2280  
 |||  
 Db 2221 AGTAAAGGATCTAGTGTGAAGGCGCAATGAGAGAAACACCTCTGAGCCAGAATCCC 2280  
 QY 2281 TTGACCTCTTCAATTTATGTAATCATATGATATGATGAGAGAAACACCTCTGAGCCAGAATCCC 2340  
 |||  
 Db 2281 TTGACCTCTTCAATTTATGTAATCATATGATATGATGAGAGAAACACCTCTGAGCCAGAATCCC 2340  
 QY 2341 GTTGTAAAGGAGTGAATGCTGATCTGTAAGTGAAGAGAAATTAAC 2389  
 |||  
 Db 2341 GTTGTAAAGGAGTGAATGCTGATCTGTAAGTGAAGAGAAATTAAC 2389

RESULT 2  
 AF291695 2541 bp DNA linear BCT 15-MAR-2001  
 LOCUS  
 DEFINITION Streptococcus pneumoniae pneumococcal histidine triad A protein  
 (phb) gene, complete cds.  
 ACCESSION AF291695  
 VERSION AF291695.1 GI:13345012  
 KEYWORDS  
 SOURCE Streptococcus pneumoniae.  
 ORGANISM Streptococcus pneumoniae.  
 Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 Streptococcus.

REFERENCE  
 AUTHORS Wismann, T. M., Heinrichs, J. H., Adamou, J. E., Erwin, A. L., Kunsch, C.,  
 Choi, G. H., Barash, S. C., Rosen, C. A., Masure, H. R., Tummen, B.,  
 Gayle, A., Brehm, Y. A., Walsh, W., Barren, P., Lathigra, R., Hanson, M.,  
 Langemann, S., Johnson, S., and Koenig, S.  
 TITLE Use of a whole genome approach to identify vaccine molecules  
 affording protection against Streptococcus pneumoniae infection  
 JOURNAL Infect. Immun. 69 (3), 1593-1598 (2001)  
 MEDLINE 2116976  
 PUBMED 11179332  
 REFERENCES  
 1 (bases 1 to 2541)  
 2 (bases 1 to 2541)  
 AUTHOR'S CHOI, G. H.  
 TITLE Direct Submission  
 JOURNAL Submitted (01-AUG-2000) Molecular Biology, Human Genome Sciences,  
 Inc., 9410 Key West Ave., Rockville, MD 20850, USA

FEATURES  
 source  
 1..2541  
 /organism="Streptococcus pneumoniae"  
 /strain="N4"  
 /db\_xref="taxon:1334"  
 91..2541  
 /gene="phbA"  
 91..2541  
 /gene="phbA"  
 /note="phbA"  
 /codon\_start=1  
 /transl\_table=11  
 /product="pneumococcal histidine triad A protein"  
 /protein\_id="AAK1915.1"  
 /db\_xref="GI:13345013"  
 /translation="MKINKKYLIVGSAALILSVSELYGQARTKENVKRVSYDOK  
 OATOKENTLPDEVSREGINAOIVIKITIDGQYVSHGHYHYRKGKVPYDAIISSE  
 LMKDPNXYKIKEDYIVNEKGVYIVKDGKYYLYLDAADAVRKEELNKRKOEHS  
 QHEGEGTPRNDGAVVALARSOGRTITDGTIFNLSDIIEDGDVIVPHGHYIYEPN  
 ELISASLEAAEAFLSGRNLSNRTYRKQSDMTSVNWPVSVNGTPTNTNSNSN  
 TNSOASQNDIDSLIKOLYKPLPSORHVESDGLVDPQAITSRTAGVAVPHGDHYH

BASE COUNT 888 a 476 c 516 g 660 t 1 others  
 ORIGIN  
 I PYSOMSELEERIALIIPLRYSNHNWPDSPPOBPPEPSPGOPAPNLKIDSN  
 SLSVQLEVRKGVGEVEEKGISRYFAKLPSEYKNLESKESVHTLTAKE  
 NVAPODEVDKAYNLTTEHAKRLFPYKNNSPOFLDLLELNESTKRELVNDL  
 LAFLEPTEHERIGKRSQIEYEDEYRIQLDKTTTSGITFDEHDIISDGDVY  
 TPRMGSHWIGKDSLSDEKRVAAOVATKREGLIPSPDADVANKANPLYSNRYK  
 GKRIPLVRLPYVEHRTVEYKGNLIIIPHDYHNITKFAFMDHTYKAPNGYLEDLF  
 ATIKRYVEHDEPERSNDGNSSEHVLKPSDEPNKFKADEPVEETPAPEVP  
 OVEKEVEAOLKEAEVYLAKVTDSSLKANETELAGLRNLTQINDNNSIMAEKVL  
 LALLKSNPSVSKEKIN"

Query Match 100.0%; Score 2388; DB 1; Length 2541;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 2389; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 TTCTTACGAGTTGGGACTGTAATCAAGCTAGACGCTTAAAGAAATTAATCGTGTTCCTA 60  
 |||  
 Db 150 TTCTTACGAGTTGGGACTGTAATCAAGCTAGACGCTTAAAGAAATTAATCGTGTTCCTA 209  
 QY 61 TATGATGGAAGAAACAGGAGCGCAAAAACGAGAAATTTGCTGCTGATGAGTTAGCAA 120  
 |||  
 Db 210 TATGATGGAAGAAACAGGAGCGCAAAAACGAGAAATTTGCTGCTGATGAGTTAGCAA 269  
 QY 121 GCGTGAAGGAATCAATGCTGAGCAAAATCGTCATCAAGATAACAGACAGGCTATGTCAC 180  
 |||  
 Db 270 GCGTGAAGGAATCAATGCTGAGCAAAATCGTCATCAAGATAACAGACAGGCTATGTCAC 329  
 QY 181 TTCAATGCGGACGACATATCATATTAACAAATGTAAGGTCCTTATGACGCTATCATCAG 240  
 |||  
 Db 330 TTCAATGCGGACGACATATCATATTAACAAATGTAAGGTCCTTATGACGCTATCATCAG 389  
 QY 241 TGAAGATTAATCAATGGAAGATCCAAACTATTAAGATTAAGATATGTAATGA 300  
 |||  
 Db 390 TGAAGATTAATCAATGGAAGATCCAAACTATTAAGATTAAGATATGTAATGA 449  
 QY 301 GGTCAAGGATGATATGATGATCAAGTATAGTGAAGATTAAGATTAAGATTAAGATGTC 360  
 |||  
 Db 450 GGTCAAGGATGATATGATGATCAAGTATAGTGAAGATTAAGATTAAGATTAAGATGTC 509  
 QY 450 GGTCAAGGATGATATGATGATCAAGTATAGTGAAGATTAAGATTAAGATTAAGATGTC 509  
 |||  
 Db 361 TGCCACGCGGATTAAGTCCGTCACAAAAGGAATCAATGACAAAACAGAGCATAG 420  
 |||  
 Db 510 TGCCACGCGGATTAAGTCCGTCACAAAAGGAATCAATGACAAAACAGAGCATAG 569  
 QY 421 TCAACATCGTGAAGGTGAAGTCCCAAGAACGATGCTGCTTGGCGACGCTTGCA 480  
 |||  
 Db 570 TCAACATCGTGAAGGTGAAGTCCCAAGAACGATGCTGCTTGGCGACGCTTGCA 629  
 QY 481 AGGACGCTATCTACAGATGATGATGATTAATGCTTCTGATATCATAGAGATAC 540  
 |||  
 Db 630 AGGACGCTATCTACAGATGATGATGATTAATGCTTCTGATATCATAGAGATAC 689  
 QY 541 TGGTATGCTTATATCGTTCCCTCATGAGATCAATTAACATTAATCTCTAAGATGAGTT 600  
 |||  
 Db 690 TGGTATGCTTATATCGTTCCCTCATGAGATCAATTAACATTAATCTCTAAGATGAGTT 749  
 QY 601 ATCAGCTAGCGAGTGGCTGCTGAGAGAGCTTCTCATCTGCTGAGAGAAATCTGTCAA 660  
 |||  
 Db 750 ATCAGCTAGCGAGTGGCTGCTGAGAGAGCTTCTCATCTGCTGAGAGAAATCTGTCAA 809  
 QY 661 TTCAAGAACCTATGCGCGACAAAATAGAGATTAACATTAACATTAACATTAACATTAACATTA 720  
 |||  
 Db 810 TTCAAGAACCTATGCGCGACAAAATAGAGATTAACATTAACATTAACATTAACATTAACATTA 869  
 QY 721 TGTAAAGCAATTCAGAACTAACAATTAACAACAAGCAACAAGCAACAAGCAACAAGCAACA 780  
 |||  
 Db 870 TGTAAAGCAATTCAGAACTAACAATTAACAACAAGCAACAAGCAACAAGCAACAAGCAACA 929  
 QY 781 AGCAAGTCAAGTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 840  
 |||  
 Db 930 AGCAAGTCAAGTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 989  
 QY 841 TCAAGCAATGTAAGATCTAGTGCCTTGCTTGTATGATCAAGCAACAATCAACAAGTCAAC 900

|||||  
Db 990 TCAACGACATGTAGAAATCTGATGGCCCTGTTGTGATCCAGCAAAATCACAAGTCAAC 1049  
QY 901 AGCTAAGAGTGTGCACTGCGCACACGAGATCAATTAACACTTCACTCCCTTACTCTCAAT 960  
Db 1050 AGCTAAGAGTGTGCACTGCGCACACGAGATCAATTAACACTTCACTCCCTTACTCTCAAT 1109  
QY 961 GTCTGAATTTGGAAGACGAATCGCTGATTTATTTCCCTTCTGTTATGCTTCAACCAATTG 1020  
Db 1110 GTCTGAATTTGGAAGACGAATCGCTGATTTATTTCCCTTCTGTTATGCTTCAACCAATTG 1169  
QY 1021 GGTACCAAGATTCAAGGCCAGAACACCAGTCCACACGCACTCCGAGACCTAGTCCAGG 1080  
Db 1170 GGTACCAAGATTCAAGGCCAGAACACCAGTCCACACGCACTCCGAGACCTAGTCCAGG 1229  
QY 1081 CCGGCAACCTGCAACCAATTTAAATAGACTCAAAATCTCTTTGGTTGTGTCAGCTGGT 1140  
Db 1230 CCGGCAACCTGCAACCAATTTAAATAGACTCAAAATCTCTTTGGTTGTGTCAGCTGGT 1289  
QY 1141 ACGAAAAGTTGGGGAAGGATATGTATTCGAAGAAAAGGCGATCTCTGTTATGCTTTGC 1200  
Db 1290 ACGAAAAGTTGGGGAAGGATATGTATTCGAAGAAAAGGCGATCTCTGTTATGCTTTGC 1349  
QY 1201 GAAAGATTACATCTGAAAACCTGTTAAAAATCTTGAAGCAAGTTATCAAAACAAGAGAG 1260  
Db 1350 GAAAGATTACATCTGAAAACCTGTTAAAAATCTTGAAGCAAGTTATCAAAACAAGAGAG 1409  
QY 1261 TGTTCACACACTTAACTTAACTTAAAGAAAATGTCCTCCGTCGACCAACAATTTTA 1320  
Db 1410 TGTTCACACACTTAACTTAACTTAAAGAAAATGTCCTCCGTCGACCAACAATTTTA 1469  
QY 1321 TGATTAAGCATATATCTGTTAACTGAGGCTCAATAAGCCTTGTGTTGNAATAAGGCTCG 1380  
Db 1470 TGATTAAGCATATATCTGTTAACTGAGGCTCAATAAGCCTTGTGTTGNAATAAGGCTCG 1529  
QY 1381 TAATTCGATTTCCAGCCTTAGACAATTTATAGAACGCTTGAGATGATGACTAA 1440  
Db 1530 TAATTCGATTTCCAGCCTTAGACAATTTATAGAACGCTTGAGATGATGACTAA 1589  
QY 1441 TAAAGAAAATTTGATGATTTATTTGGCATTCCTAGCAACCAATTAACCATCAAGAGCG 1500  
Db 1590 TAAAGAAAATTTGATGATTTATTTGGCATTCCTAGCAACCAATTAACCATCAAGAGCG 1649  
QY 1501 ACTTGGCAAAACCAATTTCTCAATTTAGATAGTGAAGAGCAAGTTGCTATCTCAATT 1560  
Db 1650 ACTTGGCAAAACCAATTTCTCAATTTAGATAGTGAAGAGCAAGTTGCTATCTCAATT 1709  
QY 1561 AGCTGATAGTATACACAGCTCAGATGCTTACATTTTGTGTAACATGATATTAATCACTGA 1620  
Db 1710 AGCTGATAGTATACACAGCTCAGATGCTTACATTTTGTGTAACATGATATTAATCACTGA 1769  
QY 1621 TGAAGAGATGATATGTATAGCGCTCATATGAGGCAATGTCAGTGGATGGAAAAAGTAG 1680  
Db 1770 TGAAGAGATGATATGTATAGCGCTCATATGAGGCAATGTCAGTGGATGGAAAAAGTAG 1829  
QY 1681 CCTTTCGATAGGAAAAAGTTGACAGCTCAAGCCTTACTAATAAGAAAAAGTATCTCAACC 1740  
Db 1830 CCTTTCGATAGGAAAAAGTTGACAGCTCAAGCCTTACTAATAAGAAAAAGTATCTCAACC 1889  
QY 1741 TCCATCTCCAGACGAGATGTTAAGCAAACTCACTGAGATAGTGCACAGCTATTTA 1800  
Db 1890 TCCATCTCCAGACGAGATGTTAAGCAAACTCACTGAGATAGTGCACAGCTATTTA 1949  
QY 1801 CAATCGTGAAGAGGGAAGAAAGCAATTCACGCTTCGATTCATATATGTTGAGCA 1860  
Db 1950 CAATCGTGAAGAGGGAAGAAAGCAATTCACGCTTCGATTCATATATGTTGAGCA 2009  
QY 1861 TACAGTTGAAGTAAAAACGTAATTTGATTTATCTCATAGAGATCATTAACATATAT 1920  
Db 2010 TACAGTTGAAGTAAAAACGTAATTTGATTTATCTCATAGAGATCATTAACATATAT 2069  
QY 1921 TAAATTTGCTTGGTTGATGATCAACATCAAAAGCTCCAATGGCTATACCTTGAAGA 1980  
|||||

Db 2070 TAAATTTGCTTGGTTGATGATCAACATCAAAAGCTCCAATGGCTATACCTTGAAGA 2129  
QY 1981 TTTGTTGGCAGATTAAGTACTAGTAGAACCCCTGACAGAACGTCACATTTCTAATGA 2040  
Db 2130 TTTGTTGGCAGATTAAGTACTAGTAGAACCCCTGACAGAACGTCACATTTCTAATGA 2189  
QY 2041 TGGATGGGCAATTCAGTATGAGCATGTGTAGGCAAGAAAGACACAGTGAAGATCCAA 2100  
Db 2190 TGGATGGGCAATTCAGTATGAGCATGTGTAGGCAAGAAAGACACAGTGAAGATCCAA 2249  
QY 2101 TAAGAACTTCAAAAGCGGATGAAGACCCAGTAGAGAAAACCTGCTGAGCAGAGTCCC 2160  
Db 2250 TAAGAACTTCAAAAGCGGATGAAGACCCAGTAGAGAAAACCTGCTGAGCAGAGTCCC 2309  
QY 2161 TCAAGTAGACGTGAAGAAATAGAGCCCAATCAAAAGACGAAGTTTGTGTCGAA 2220  
Db 2310 TCAAGTAGACGTGAAGAAATAGAGCCCAATCAAAAGACGAAGTTTGTGTCGAA 2269  
QY 2221 AGTACGGATTCTAGTCTGAAGCCCAATGCAACAGAAAACCTGAGCTGTTACGAATAA 2280  
Db 2370 AGTACGGATTCTAGTCTGAAGCCCAATGCAACAGAAAACCTGAGCTGTTACGAATAA 2429  
QY 2281 TTTGACTCTTCAAAATTAAGTATACCAATAGTATCATGCGCAAGCAAAAAATTAATTGC 2340  
Db 2430 TTTGACTCTTCAAAATTAAGTATACCAATAGTATCATGCGCAAGCAAAAAATTAATTGC 2489  
QY 2341 GTTGTAAAGAAAGTATATCTCATCTGTAAGTAAAGAAAATAAAC 2389  
Db 2490 GTTGTAAAGAAAGTATATCTCATCTGTAAGTAAAGAAAATAAAC 2538  
|||||

RESULT 3  
BD003774  
LOCUS BD003774 8195 bp DNA linear PART 31-JAN-2002  
DEFINITION Polynucleotide of Streptococcus pneumoniae and sequence.  
ACCESSION BD003774.1 GI:18631735  
VERSION JP 2001501833-A/94.  
KEYWORDS unclassified.  
SOURCE unclassified.  
ORGANISM unclassified.  
REFERENCE 1 (bases 1 to 8195)  
AUTHORS Kunsch, C.A., Choi, G.H., Dillon, P.J., Rosen, C.A., Bara, S.C.,  
Fannon, M. and Dougherty, B.A.  
TITLE Polynucleotide of Streptococcus pneumoniae and sequence  
JOURNAL Patent: JP 2001501833-A 94 13-FEB-2001;  
HUMAN GENOME SCIENCES INC  
COMMENT OS unclassified  
PN JP 2001501833-A/94  
PD 13-FEB-2001  
PE 30-OCT-1997 JP 1998520718  
PR 31-OCT-1996 US 60/029960  
PI CHARLES A KUNSCH, GIL H CHOI, PATRICK J DILLON, CRAIG A ROSEN, PI  
STEVEN C BARASH,  
PI MICHAEL FANNON, BRIAN A DOUGHERTY  
PC C12N15/09, A01K67/027, C07K14/315, C07K16/12, C12N1/15, C12N1/19,  
PC C12N1/21,  
PC C12N5/10, C12P21/02, C12Q1/68, G06F17/30, C12N15/00, C12N5/00, PC  
G06F15/40  
CC Strandedness: Double;  
CC Topology: Linear;  
FH key Location/Qualifiers  
FT source 1..8195  
LOCATION/Qualifiers  
1..8195  
Location/Qualifiers  
/organism='unclassified'  
/db\_xref='taxon:32644'  
BASE COUNT 2688 a 1622 c 1777 g 2105 t 3 others  
ORIGIN  
Query Match 100.0%; Score 2388; DB 6; Length 8195;  
Best Local Similarity 100.0%; Pred. No. 0;

Matches 2388: Conservative 0; Mismatches 1; Indels 0; Gaps 0;				
QY	1	TTCTTAGAGTTGGAGCTGTATCAAGCTAGAACGTTAAGAAAAATATGTTGCTTCCFA	60	
Db	3053	TTCTTAGAGTTGGAGCTGTATCAAGCTAGAACGTTAAGAAAAATATGTTGCTTCCFA	3112	
QY	61	TATAGATGAAAAACAGCAGCGCAAAAAAGGAGAAATTTGACTCTGATGAGGTTACAA	120	
Db	3113	TATAGATGAAAAACAGCAGCGCAAAAAAGGAGAAATTTGACTCTGATGAGGTTACAA	3172	
QY	121	GGCTGAAGAAATCAATCTGAGCAAAATCGTCATCAAGATTAACAGACCAAGCTATGTCAC	180	
Db	3173	GGCTGAAGAAATCAATCTGAGCAAAATCGTCATCAAGATTAACAGACCAAGCTATGTCAC	3232	
QY	181	TTACATGCGGACCATATCATTTATTTCAATGTAAGTTCCTTATGAGCTATCATCAG	240	
Db	3233	TTACATGCGGACCATATCATTTATTTCAATGTAAGTTCCTTATGAGCTATCATCAG	3292	
QY	241	TGAAGAAATTAATCTATGAAAGATCCAAACTATAGCTAAAAAGATGAGATATTTGTTAATGA	300	
Db	3293	TGAAGAAATTAATCTATGAAAGATCCAAACTATAGCTAAAAAGATGAGATATTTGTTAATGA	3352	
QY	301	GGTCAAGGGGAGATATGTTTTCAAAGTAGATGAAATTAATGTTTACTATGTTTAAAGATGC	360	
Db	3353	GGTCAAGGGGAGATATGTTTTCAAAGTAGATGAAATTAATGTTTACTATGTTTAAAGATGC	3412	
QY	361	TGCCACGCGGATTAACGTCCGTACAAAGAGAAATCAATCGACAAAACAGACATAG	420	
Db	3413	TGCCACGCGGATTAACGTCCGTACAAAGAGAAATCAATCGACAAAACAGACATAG	3472	
QY	421	TCAACATCGTGAAGGTGAACTCCAGAAAAGATGTGCTGTGCTTGGCAAGCTTGCA	480	
Db	3473	TCAACATCGTGAAGGTGAACTCCAGAAAAGATGTGCTGTGCTTGGCAAGCTTGCA	3532	
QY	481	AGGAGCTATCTACAGATAGATGTTTATCTTAATGCTTGCATATTCATAGAGATAC	540	
Db	3533	AGGAGCTATCTACAGATAGATGTTTATCTTAATGCTTGCATATTCATAGAGATAC	3592	
QY	541	TGGTATGCTTATATCGTTCCTCATGAGATCAATTACATTACCTTCTAAGATGAGTT	600	
Db	3593	TGGTATGCTTATATCGTTCCTCATGAGATCAATTACATTACCTTCTAAGATGAGTT	3652	
QY	601	ATGAGCTAGCGAGTTGGCTGCTCGAGAAAGCTTCTATCTGTGTCGAGAAATCTGTCAA	660	
Db	3653	ATGAGCTAGCGAGTTGGCTGCTCGAGAAAGCTTCTATCTGTGTCGAGAAATCTGTCAA	3712	
QY	661	TTCAAGAACTATTCGCCGAGAAAATATACGATTAACACTTTCAAGAACTAGGAGCTTC	720	
Db	3713	TTCAAGAACTATTCGCCGAGAAAATATACGATTAACACTTTCAAGAACTAGGAGCTTC	3772	
QY	721	TGTAAAGATCCAGAACTCAAAATACTAACAAGCAAGCAAGCAAGCAACTAACACTCA	780	
Db	3773	TGTAAAGATCCAGAACTCAAAATACTAACAAGCAAGCAAGCAAGCAACTAACACTCA	3832	
QY	781	AGCAAGTCAAGTAATGACATTTAGTCTTTGAAAACAGCTCTACAACTGCTTTGAG	840	
Db	3833	AGCAAGTCAAGTAATGACATTTAGTCTTTGAAAACAGCTCTACAACTGCTTTGAG	3892	
QY	841	TCAACGACATGTAGATCTATGAGCTTGTGCTTGTGATCCGACAAATCAACAGTGCAC	900	
Db	3893	TCAACGACATGTAGATCTATGAGCTTGTGCTTGTGATCCGACAAATCAACAGTGCAC	3952	
QY	901	AGCTAGAGTGTTCAGATGTCACACGAGATCAATTACACTTCATCCCTTACTCTCAAT	960	
Db	3953	AGCTAGAGTGTTCAGATGTCACACGAGATCAATTACACTTCATCCCTTACTCTCAAT	4012	
QY	961	GCTGTAATTTGGAAGAAAGAAATCGCTGTATTTATTTCCCTTGTATTCGTTCAACATTTG	1020	
Db	4013	GCTGTAATTTGGAAGAAAGAAATCGCTGTATTTATTTCCCTTGTATTCGTTCAACATTTG	4072	
QY	1021	GGTAAAGATTTCAAGGCAACAAACAACTCCACAAACGACATCTCGAAGCTAGTCCAG	1080	
Db	4073	GGTAAAGATTTCAAGGCAACAAACAACTCCACAAACGACATCTCGAAGCTAGTCCAG	4132	

---

QY	1081	CCGCAACCTGCACCAATCTTAAATAGACTCAAAATCTTCTTGGTATGACGTGCT	1140	
Db	4133	CCGCAACCTGCACCAATCTTAAATAGACTCAAAATCTTCTTGGTATGACGTGCT	4192	
QY	1141	ACGAAAAGTTGGGAGATATGATTTGCAAGAAAAGGAGATCTCTGTTATGCTTTGC	1200	
Db	4193	ACGAAAAGTTGGGAGATATGATTTGCAAGAAAAGGAGATCTCTGTTATGCTTTGC	4252	
QY	1201	GAAAGATTTTACATCTGAAATCTTAAATCTTAAATCTTAAATCTTAAATCTTAAATCTTAA	1260	
Db	4253	GAAAGATTTTACATCTGAAATCTTAAATCTTAAATCTTAAATCTTAAATCTTAAATCTTAA	4312	
QY	1261	TGTTTACACACTTAACTCTGCTTAAAGAAAATGTTGCTCTCTGTCGACCAAGATTTTA	1320	
Db	4313	TGTTTACACACTTAACTCTGCTTAAAGAAAATGTTGCTCTCTGTCGACCAAGATTTTA	4372	
QY	1321	TGATTAAGCATATATCTGTTTAACTGAGGCTCATTAAGCCTTGTGTTGMAATAAGGGTGC	1380	
Db	4373	TGATTAAGCATATATCTGTTTAACTGAGGCTCATTAAGCCTTGTGTTGMAATAAGGGTGC	4432	
QY	1381	TAAATCTGATTTCCAAAGCTTAAAGCAATTTATTAAGACCTTGAATGATTAATGACTAA	1440	
Db	4433	TAAATCTGATTTCCAAAGCTTAAAGCAATTTATTAAGACCTTGAATGATTAATGACTAA	4492	
QY	1441	TAAAGAAAATTTGGTAGATATTTATGCAATTCCTAGACCAATTAACCATCCAGAGCG	1500	
Db	4493	TAAAGAAAATTTGGTAGATATTTATGCAATTCCTAGACCAATTAACCATCCAGAGCG	4552	
QY	1501	ACTTGGCAAAACCAATTTCTCAAAATGAGTATCTGAAGACGAAGTTGATGCTCAATT	1560	
Db	4553	ACTTGGCAAAACCAATTTCTCAAAATGAGTATCTGAAGACGAAGTTGATGCTCAATT	4612	
QY	1561	AGCTGATTAAGTATCAACGTCAGATGTTTACATTTTGATGAACATGATTAATCACTGA	1620	
Db	4613	AGCTGATTAAGTATCAACGTCAGATGTTTACATTTTGATGAACATGATTAATCACTGA	4672	
QY	1621	TGAAGAGATGATATGATTAACGCTCATATGAGGCTAGTCACTGATTTGAAAAGATAG	1680	
Db	4673	TGAAGAGATGATATGATTAACGCTCATATGAGGCTAGTCACTGATTTGAAAAGATAG	4732	
QY	1681	CTTTTCTGATTAAGGAAAAGTTGCAAGCTCAAGCTATTAAGAAAAGATATCTTACC	1740	
Db	4733	CTTTTCTGATTAAGGAAAAGTTGCAAGCTCAAGCTATTAAGAAAAGATATCTTACC	4792	
QY	1741	TCCATCTCCAGAGCAATGTTAAAGCAATTCCAACTGGAGATAGTCCAGCAGCTATTTA	1800	
Db	4793	TCCATCTCCAGAGCAATGTTAAAGCAATTCCAACTGGAGATAGTCCAGCAGCTATTTA	4852	
QY	1801	CAATCGTGAAGGAGGAAAAGCAATTCACATGCTTCGATTCGATTAATGTTGAGCA	1860	
Db	4853	CAATCGTGAAGGAGGAAAAGCAATTCACATGCTTCGATTCGATTAATGTTGAGCA	4912	
QY	1861	TACAGTTGAGGTTAAAAACGTAATTTGATTAATTCCTCATTAAGATCATTAACATAAT	1920	
Db	4913	TACAGTTGAGGTTAAAAACGTAATTTGATTAATTCCTCATTAAGATCATTAACATAAT	4972	
QY	1921	TAAATTTGCTTGGTTTATATATACACATACAAAGCTCCAAATGCTATATCTTGGAGA	1980	
Db	4973	TAAATTTGCTTGGTTTATATATACACATACAAAGCTCCAAATGCTATATCTTGGAGA	5032	
QY	1981	TTTGTTCGACGATTTAAGTACTAGTAGAACAACCCGAGAGAGTCCACATTTCTAATGA	2040	
Db	5033	TTTGTTCGACGATTTAAGTACTAGTAGAACAACCCGAGAGAGTCCACATTTCTAATGA	5092	
QY	2041	TGATGGGGAATTCAGTAGAGCATGTTAGGCCAAGAAAGACACAGTAAAGATCCAAA	2100	
Db	5093	TGATGGGGAATTCAGTAGAGCATGTTAGGCCAAGAAAGACACAGTAAAGATCCAAA	5152	
QY	2101	TAAAGACTTTCAAAAGCGATGAAGAGCCAGTAGAGAAAACCTGCTGAGCAAGATGCC	2160	
Db	5153	TAAAGACTTTCAAAAGCGATGAAGAGCCAGTAGAGAAAACCTGCTGAGCAAGATGCC	5212	

QY 2161 TCAAGTAGACACTGAAAAGTAGAGCCCACTCAAGACAGAAAGTTTGGTCGGAA 2220  
|||||  
Db 5213 TCAAGTAGAGACTGAAAAGTAGAGCCCACTCAAGACAGAAAGTTTGGTCGGAA 5272  
QY 2221 AGTAAGGATTTAGTCTGTAAGCAATGCAAGAACTAGCTGGTTAGCAAAATA 2280  
|||||  
Db 5273 AGTAAGGATTTAGTCTGTAAGCAATGCAAGAACTAGCTGGTTAGCAAAATA 5332  
QY 2281 TTTGACTCTTCAATTTATGATTAACATATGATGAGCAAGCAAGAAATTTACTTGC 2340  
|||||  
Db 5333 TTTGACTCTTCAATTTATGATTAACATATGATGAGCAAGCAAGAAATTTACTTGC 5392  
QY 2341 GTTGTAAAAAGAGTAATCTTCATCTGTAAGTAAGCAAAAAATAAAC 2389  
|||||  
Db 5393 GTTGTAAAAAGAGTAATCTTCATCTGTAAGTAAGCAAAAAATAAAC 5441

RESULT 4  
AE007418/c 10256 bp DNA linear BCF 31-AUG-2001  
LOCUS Streptococcus pneumoniae TIGR4 section 101 of 194 of the complete  
DEFINITION genome.  
ACCESSION AE007418 AE005672  
VERSION AE007418.1 GI:14972649  
KEYWORDS  
SOURCE Streptococcus pneumoniae TIGR4.  
ORGANISM Streptococcus pneumoniae TIGR4  
Bacteria: Firmicutes; Lactobacillales; Streptococcaceae;  
Streptococcus.  
REFERENCE 1 (bases 1 to 10256)  
Tetzelin,H., Nelson,K.E., Paulsen,I.T., Eisen,J.A., Read,T.D.,  
Peterson,S., Heidelberg,J., Deboy,R.T., Haft,D.H., Dodson,R.J.,  
Durrin,A.S., Gwinn,M., Kolonay,J.F., Nelson,W.C., Peterson,J.D.,  
Umayam,L.A., White,O., Salzberg,S.L., Lewis,M.R., Radune,D.,  
Holtzapple,E., Khouri,H., Wolf,A.M., Uitterback,T.R., Hansen,C.L.,  
McDonald,L.A., Feldblyum,T.V., Angiuoli,S., Dickinson,T.,  
Hickey,E.K., Holt,I.E., Loftus,B.J., Yang,F., Smith,H.O.,  
Venter,J.C., Dougherty,B.A., Morrison,D.A., Hollingshead,S.K. and  
Fraser,C.M.  
TITLE Complete genome sequence of a virulent isolate of Streptococcus  
pneumoniae  
JOURNAL Science 293 (5529), 498-506 (2001)  
MEDLINE 21357209  
PUBMED 11463916  
REFERENCE 2 (bases 1 to 10256)  
Tetzelin,H., Nelson,K.E., Paulsen,I.T., Eisen,J.A., Read,T.D.,  
Peterson,S., Heidelberg,J., Deboy,R.T., Haft,D.H., Dodson,R.J.,  
Durrin,A.S., Gwinn,M., Kolonay,J.F., Nelson,W.C., Peterson,J.D.,  
Umayam,L.A., White,O., Lewis,M.R., Radune,D., Holtzapple,E.,  
Khouri,H., Wolf,A.M., Uitterback,T.R., Hansen,C.L., McDonald,L.A.,  
Feldblyum,T.V., Angiuoli,S., Gesuwan,P., Hickey,E.K., Holt,I.E.,  
Loftus,B.J., Ujwal,M.L., Yang,F., Smith,H.O., Venter,J.C.,  
Dougherty,B.A., Morrison,D.A., Hollingshead,S.K. and Fraser,C.M.  
TITLE Direct Submission  
JOURNAL Submitted (29-JUN-2001) The Institute for Genomic Research, 9712  
Medical Center Dr, Rockville, MD 20850, USA  
FEATURES  
source 1..10256  
/organism="Streptococcus pneumoniae TIGR4"  
/strain="TIGR4"  
/db\_xref="taxon:170187"  
/gene="SP1170"  
complement(100..702)  
complement(100..702)  
/gene="SP1170"  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAK75279.1"  
/db\_xref="GI:14972650"  
/translation="MSEVDPRNAVYEFPSDTCOLANSYGLFEPDKKNSCLLIF  
TWKSPSLVKEGDYIGRDSQVDNLNVIGNIIFPNYLTNRKYSLNMRNCGMGEPPHDF

DIYLDHAVKAYAEQVNNIKKEYPLKRALIHQENALYRFFSNPDFLEKNLYKTWO  
VSKETPFSEMDFMFKNIKSEKIIIFERGSFMLNDLKSNNK"  
complement(760..1473)  
/gene="SP1171"  
complement(760..1473)  
/gene="SP1171"  
/note="identified by match to Pfam protein family HM  
PF00702"  
/codon\_start=1  
/transl\_table=11  
/product="hydrolase, haloacid dehalogenase-like family"  
/protein\_id="AAK75280.1"  
/db\_xref="GI:14972651"  
/translation="MEYKFLFEDLDLDPDAEDVALTQLKEEGVADIQAYKDY  
VPMKALMKDELKISKOEIVNTRFSLFAHGOEGSFLAORFOYLAQOGOTLS  
GAHDLIDLIERDVNLVAATNGITAIOTGRILAOSGLAFYNOVFIISDLOLQKPDALF  
YKIKOQIANGFESKEKTIAMGDSLTADIGGNNAGIDITTYNPHHLENTHQAOPTVEY  
STQDILLDCDKNLEKITY"  
complement(1724..1843)  
/gene="SP1172"  
complement(1724..1843)  
/gene="SP1172"  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAK75281.1"  
/db\_xref="GI:14972652"  
/translation="MLIVAPFYFNSEITVFLNCESITDCDGAILEYTERKN"  
complement(1933..2307)  
/gene="SP1173"  
complement(1933..2307)  
/gene="SP1173"  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAK75282.1"  
/db\_xref="GI:14972653"  
/translation="MESIGLIVTSKHAEGVVELISKVAKDPTIYVREGGIG  
TSFEQVDVYSENPAIDTLAFEDLSAKNNLKVYDFSDKSLIINRYIVIEGAYNAAA  
LLQAGAEISVIGTQALELEINK"  
complement(2729..5188)  
/gene="SP1174"  
complement(2729..5188)  
/gene="SP1174"  
/note="similar to GP:4249624; identified by sequence  
similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved domain protein"  
/protein\_id="AAK75283.1"  
/db\_xref="GI:14972654"  
/translation="MKINKYLAGVAVLAISVCSYELGRIOAGODKESNRVAYIDG  
DQAGKAPNLTPEVSKREGINAEQIVIKITDQGVTSGHYVNGKVPYDAIIS  
ELMKDPYOLKSDIVNEIKGVYIKNGYVYLAKPAADNRITREKEETKQOER  
SHNHSRADNVAARAGRYTTDDGYIFNPSADITDEGDVAVTYPMTHSHWKKSLS  
SASELAADAVWNGKGRSPSSSSSYNANPQRISEHNHNLVPTVYONONGENISL  
LRELYAKLSERHVESDILIDPRAOITSRTARGAVVPNGNHYHETPYQMSLEKRIA  
RLIPLRTSNHWPDSRPEEPSPOETPEPSPOAPSNPIDELKVAEVAKGVGGVY  
FEENGVSRYIPAKDISAETAGIDSKLAKQSLSKLTGKTDLPSDSOREFYNNAYDL  
LARIHQDLIDNKGQVDFEALDNLEIRLKVSDSKVLEVDILAFAPIRPERKGP  
NAQIYVDEIDQVAKLAKYTTEDGYIFDPDITDEGDVAVTYPMTHSHWKKSLS  
EAERAAQAVAKREGLTPSPDHDOSGTEAKGAATYINRYAKAKKVPDLDMPNVLOY  
TVEYKNGSLIIPHYDHYNIKFEWFDDELTYAPKGYTLEDLATYKYVVEHPNPHS  
DNREGNASDHYORNKNGADTNQTEKPESEKPOTEKPEETPREKPOSEKPEKPT  
EEPESEPESEPEQVETEKREABDLGKIDPILIKSNARETTLGLKNNLLFGT  
QDNNTIIMAEKLLALEKSK"  
complement(5346..7754)  
/gene="SP1175"  
complement(5346..7754)  
/gene="SP1175"  
/note="similar to GP:4249624; identified by sequence

```
similarity: putative"
/codon_start=1
/transl_table=11
/product="conserved domain protein"
/protein_id="AAK75284.1"
/db_xref="GI:14972655"
/translacion="MILSVCSYELHLYQARTVKNRVSIDKQATOKTENLTPDEV
SKREGINADIVIKITIDOGVYTHSHGHYHNGKVPYD1ISELLMKPNYLKDED
IVNEVGVYIKDKKYIYLAKDAHADVRKEEDINPQOEHSOREGTPENDAV
AARSGRGTDDGYIFENASDIIEDTGDVIVPHGHYHHPENELASRSLAAEFL
SGRGLNSRTRYRONSQDMSRTNWPYSNPECTTNTNNSNTNSQASQSDIDSL
LKQILPLPSQRHVESDGLVFPDPAQITTSRARGVAIPHGDHFIYQSWSLEERIA
RIIPLEYSRSHWVPDSRPEQSPQTPPEPSPPQAPNPKIDSNSSLSQVLEKVEG
VYEEKGISRYVFAKDLPSYTKNLESKLSKQESVSHLTAKENVAAPRQVLYEKAY
NLTLEAKALFEKNGRNSDFOALDKLELRNDESTEKELVDLLAFAPITPERG
KPSQLEYEDEVRIADLKTYTSDGYIFEDHITISDEGDVATPMPHSHHIGKDS
LSDKEVAQAYTEKGGILPSPDADYKANPCGDSAAIYNRYKGRKIPLYLPMV
EHRVEYKNGNLIIPKDHYNHNFANFDDHTYKAPNGITLEDLPAITIKTYVEHPDEP
HSDNGNASEHYLGRKHSSEDPNKNFKADEEVEETPAPEVPOVETKREVERQLEA
EVLAKVYDSSLKANATELTLGRNMLTIQIMDNNSIMEAEKLIALLLKSNPSYSK
EKIN"
complement(8049. .9782)
/gene="SP1176"
complement(8049. .9782)
/gene="SP1176"
/transl_table=11
/codon_start=1
/product="phosphoenolpyruvate-protein phosphotransferase"
/protein_id="AAK75285.1"
/db_xref="GI:14972656"
/translacion="MTEMLKGIASDGVAVAKAYLLVOPDLSEFTITVEDTNAEARL
DIALQASODELSVIREKAVGTGEEBAQVFAHVLADPEMSIQIKETIRAKVAAE
AGLEKTYDMETITEGEMEDNPYMOERADIRDTYKLVANLLKIKPNPASIIEEIV
IAHDLPSDPAOLDKNKFAVYNIGRSHSIAKRLTETIAVGTNNTETLYKGD
IIAVNGITGEVINNPDEQAERKAGEYARAKAEMALLKDAQITYTAGKHELAN
IGPKDVEYNNNGNAEAVGLRTEFLYMSQDEPTDEQTEAKVLEENNGPVVVR
TMDIGDKELPYFDMPEHNPFLGFRALRISSETGDANFRIOIRALLRVSNGQIRI
MEPMVALKEFRAAKAVPDEKANKLAEQAVANDVOGIMEIPAAAMADQFAKEV
DEFSISTNDLQITMAADRNNEQVSYLYQPYNSIIRLINNVKKAHABGKMGKE
MAGDOQAVPLVGMGLDEFSMSATSVLRSLMKKIDTAKMEYANARALTCESTMEV
LELOKRYVNF"
complement(9788. .10051)
/gene="SP1177"
complement(9788. .10051)
/gene="SP1177"
/codon_start=1
/transl_table=11
/product="phosphocarrier protein Hpr-"
/protein_id="AAK75286.1"
/db_xref="GI:14972657"
/translacion="NASKDFHVVAFTGIAHAPATLLVOTASKFASDITLEYKGSVNL
KSIIMGVSGVGGADVTISABGADADDAIAISETMEKEGLA"
BASE COUNT      2842 a      2089 c      1850 g      3475 t
ORIGIN
Query Match      100.0%; Score 2388; DB 1; Length 10256;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2388; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 TTCTTAGAGATTGGGACTGTATCAAGCTAGAACGGTTAAGGAAATATATGCTGTTTCTTA 60
|||||
Db 7737 TTCTTAGAGATTGGGACTGTATCAAGCTAGAACGGTTAAGGAAATATATGCTGTTTCTTA 7678
QY 61 TATAGATGAAAAACACGACGCAAAAAAGAGAAATTTGATCCTCCTATGAGGTTACCA 120
|||||
Db 7677 TATAGATGAAAAACACGACGCAAAAAAGAGAAATTTGATCCTCCTATGAGGTTACCA 7618
QY 121 GCGTGAAGAAATCAATGCTGACCAATGCTATCAAGATACAGACCAAGGCTATGTCCAC 180
|||||
Db 7617 GCGTGAAGAAATCAATGCTGACCAATGCTATCAAGATACAGACCAAGGCTATGTCCAC 7558
QY 181 TTACATGCGGACCATATATTTATGCAATGGTAGGTTCTTTATGACGCTATCATCAG 240
|||||
```

```
Db 7557 TTACATGCGGACCATATATTTATGCAATGGTAGGTTCTTTATGACGCTATCATCAG 7498
QY 241 TGAAGATTACTCATGAAGATCCAAACTATTAAGCTAAAGATGAGATATTTGTTATGA 300
|||||
Db 7497 TGAAGATTACTCATGAAGATCCAAACTATTAAGCTAAAGATGAGATATTTGTTATGA 7438
QY 301 GGTCAAGGTTGATATGTTATCAAGTAGATGGAATAATCTATGTTTACCTTAAGATGC 360
|||||
Db 7437 GGTCAAGGTTGATATGTTATCAAGTAGATGGAATAATCTATGTTTACCTTAAGATGC 7378
QY 361 TGCCCAAGCGGATTAAGTCCGCTACCAAAAGAGAAATCAATCGACAAAACAGACATAG 420
|||||
Db 7377 TGCCCAAGCGGATTAAGTCCGCTACCAAAAGAGAAATCAATCGACAAAACAGACATAG 7318
QY 421 TCAACATCGTGAAGTGAAGTCCCAAGAAAGATGGTGGTCCCTGGACAGCTTGCA 480
|||||
Db 7317 TCAACATCGTGAAGTGAAGTCCCAAGAAAGATGGTGGTCCCTGGACAGCTTGCA 7258
QY 481 AGACGCTATACATACAGATGATGTTATCTTTAATGCTTGTATATCATAGAGATAC 540
|||||
Db 7257 AGACGCTATACATACAGATGATGTTATCTTTAATGCTTGTATATCATAGAGATAC 7198
QY 541 TGGTGATGCTTATATGCTTCTCATGGAGATCATTTACCATTTACTTCTTAAGATGATT 600
|||||
Db 7197 TGGTGATGCTTATATGCTTCTCATGGAGATCATTTACCATTTACTTCTTAAGATGATT 7138
QY 601 ATCAGCTAGGAGATTGGCTGCTGTCAGAACGCTTCCTATGCTGTCAGGAATGCTGCAA 660
|||||
Db 7137 ATCAGCTAGGAGATTGGCTGCTGTCAGAACGCTTCCTATGCTGTCAGGAATGCTGCAA 7078
QY 661 TTCAGAACCTATGCGCCGACAAAATAGCATTAACCTTCAAGAAACAACCTGGTACCTTC 720
|||||
Db 7077 TTCAGAACCTATGCGCCGACAAAATAGCATTAACCTTCAAGAAACAACCTGGTACCTTC 7018
QY 721 TGTAGCAATTCAGAACTCAATTAACACAGACACACAGACAACTAACAGTCA 780
|||||
Db 7017 TGTAGCAATTCAGAACTCAATTAACACAGACACACAGACAACTAACAGTCA 6958
QY 781 AGCAGTCAAGTAATGACATGATGCTGCTGTAACACGCTGACAACTGCTTGGAG 840
|||||
Db 6957 AGCAGTCAAGTAATGACATGATGCTGCTGTAACACGCTGACAACTGCTTGGAG 6898
QY 841 TCAACGACATGTAGAAATCTGATGCGCTTGTCTTGATCCAGCAAAATCAAGTCGAC 900
|||||
Db 6897 TCAACGACATGTAGAAATCTGATGCGCTTGTCTTGATCCAGCAAAATCAAGTCGAC 6838
QY 901 AGCTAGAGGTGTGCACTGCCACAGGAGATCATTTACCACTTCACTCTCTCAAT 960
|||||
Db 6837 AGCTAGAGGTGTGCACTGCCACAGGAGATCATTTACCACTTCACTCTCTCAAT 6778
QY 961 GTCGATTTGGAAGAACGATGCTGATTTATCCCTCGTTATGCTTCAACCAATG 1020
|||||
Db 6777 GTCGATTTGGAAGAACGATGCTGATTTATCCCTCGTTATGCTTCAACCAATG 6718
QY 1021 GGTACAGATTCAAGGCGCAGAACACCAAGTCCACACGACCTCGGAACCTATGTCAGG 1080
|||||
Db 6717 GGTACAGATTCAAGGCGCAGAACACCAAGTCCACACGACCTCGGAACCTATGTCAGG 6658
QY 1081 CCGGCAACCTGACCAAAATCTTAAATATGACTCAAAATTTCTTGGTTAGTCACTGGT 1140
|||||
Db 6657 CCGGCAACCTGACCAAAATCTTAAATATGACTCAAAATTTCTTGGTTAGTCACTGGT 6598
QY 1141 ACGAAAAGTTGGGGAAGATATGATTCGAAGAAAAGGGGATGCTGCTATGCTTGGC 1200
|||||
Db 6597 ACGAAAAGTTGGGGAAGATATGATTCGAAGAAAAGGGGATGCTGCTATGCTTGGC 6538
QY 1201 GAAAGATTACATCTGAAAATCTTGAAGAACAGTTATCAAAACAGAGAG 1260
|||||
Db 6537 GAAAGATTACATCTGAAAATCTTGAAGAACAGTTATCAAAACAGAGAG 6478
QY 1261 TGTTCACACACTTTAATCTGTAAGAAATGTTGCTCTGTCGACCAAGAAATTTTA 1320
|||||
Db 6477 TGTTCACACACTTTAATCTGTAAGAAATGTTGCTCTGTCGACCAAGAAATTTTA 6418
```



QY	1321	TGTAAGCAATATATCTGTAACTAGAGGCTCTAAAGCCTGTTGNAATTAAGGGTCG	1380
Db	6417	TGATTAAGCAATATATCTGTAACTAGAGGCTCTAAAGCCTGTTGNAATTAAGGGTCG	6358
QY	1381	TAATTCGATTTCCAAACCTTTAGACAAATTAATAGAACCTTGATGATGAACTGACTAA	1440
Db	6357	TAATTCGATTTCCAAACCTTTAGACAAATTAATAGAACCTTGATGATGAACTGACTAA	6298
QY	1441	TAAAGAAAATTGGTAGATCATTTATTTGGCATTCTAGCACAATTTACCATCCAGACG	1500
Db	6297	TAAAGAAAATTGGTAGATCATTTATTTGGCATTCTAGCACAATTTACCATCCAGACG	6238
QY	1501	ACTTGCGAAACCAAAATTCGAAATTTAGATACATGTAAGACGAAGTTGATTTGCTCAAT	1560
Db	6237	ACTTGCGAAACCAAAATTCGAAATTTAGATACATGTAAGACGAAGTTGATTTGCTCAAT	6178
QY	1561	AGCTGATTAAGTATACAACGTCAGATGGTTACATTTTGTATGATCATGATATATCTGTGA	1620
Db	6117	TGAAGGAGATGCAATATGTAAGCGCTCATATATGGGCGCATCTCATCTGATTTGGAAAAATAG	1680
QY	1681	CGTTTCGATTAAGGAAAAAGTTTGCAGCTCAGAGCTATACTTAAGAAAAAGGTATCCTACC	1740
Db	6057	CGTTTCGATTAAGGAAAAAGTTTGCAGCTCAGAGCTATACTTAAGAAAAAGGTATCCTACC	5998
QY	1741	TCCCATTCACAGCCAGATGTTTAAACCAATTCACACTGAGATTAATGACAGACGCTATTTA	1800
Db	5997	TCCCATTCACAGCCAGATGTTTAAACCAATTCACACTGAGATTAATGACAGACGCTATTTA	5938
QY	1801	CAATCGTGTGAAGGGGAAAAAGCAATTCACGTCCTCACTCCACTTCATATATGTTGAGCA	1860
Db	5937	CAATCGTGTGAAGGGGAAAAAGCAATTCACGTCCTCACTCCACTTCATATATGTTGAGCA	5878
QY	1861	TACAGTTGAGGTTTAAAAACGGTAATTTGATTAATCTCTCAATAAGATCATTAACATATAT	1920
Db	5877	TACAGTTGAGGTTTAAAAACGGTAATTTGATTAATCTCTCAATAAGATCATTAACATATAT	5818
QY	1921	TAAATTTGCTGGTTGTTATGATATACACATTAACAAGCTCCAAATGGCTTTACTCTTGAAAG	1980
Db	5817	TAAATTTGCTGGTTGTTATGATATACACATTAACAAGCTCCAAATGGCTTTACTCTTGAAAG	5758
QY	1981	TTTGTTCGACGATTAATAGTACTAGCTAGAACACCTCTGACGAGCTCCACATTCATATGA	2040
Db	5757	TTTGTTCGACGATTAATAGTACTAGCTAGAACACCTCTGACGAGCTCCACATTCATATGA	5698
QY	2041	TGGATGGGGCAATCCCGTAGAGCATGTTGTAGGCAAGAAAGACCACTGSAAGATCCAAA	2100
Db	5697	TGGATGGGGCAATCCCGTAGAGCATGTTGTAGGCAAGAAAGACCACTGSAAGATCCAAA	5638
QY	2101	TAAAGACTTCAAAAGCGGATGAAGAGCCAGTAGAGCAACCTGCTGAGCCAGAAAGTCCC	2160
Db	5637	TAAAGACTTCAAAAGCGGATGAAGAGCCAGTAGAGCAACCTGCTGAGCCAGAAAGTCCC	5578
QY	2161	TCAGATGAGACGTGAAAAAGTAAACCCCACTCAAAAGACGAGAAAGTTTTCGTTGGGAA	2220
Db	5577	TCAGATGAGACGTGAAAAAGTAAACCCCACTCAAAAGACGAGAAAGTTTTCGTTGGGAA	5518
QY	2221	AGTAACGATTTAGTCTGTAAGACCAATCAACAGACAAACTCTAGCTGGGTTTACGAAATTA	2280
Db	5517	AGTAACGATTTAGTCTGTAAGACCAATCAACAGACAAACTCTAGCTGGGTTTACGAAATTA	5458
QY	2281	TTTGACGCTTCAAAATTAATGATTAACAATAGTATCATGTGCGAAGCAGAAAAATTAAGTTC	2340
Db	5457	TTTGACGCTTCAAAATTAATGATTAACAATAGTATCATGTGCGAAGCAGAAAAATTAAGTTC	5398
QY	2341	GTTGCTTAAAGGAAGTAATCCTTCATCTGTAAGTAAAGAAAAAATAAAC	2389
Db	5397	GTTGCTTAAAGGAAGTAATCCTTCATCTGTAAGTAAAGAAAAAATAAAC	5349

[illegible]

gene  
/organism="Streptococcus pneumoniae R6"  
/strain="R6"  
/db\_xref="taxon:171101"  
complement(91..738)  
/gene="spri056"  
complement(91..738)  
/gene="spri056"  
/codon\_start=1  
/transl\_table=11  
/label=spri056  
/product="Hypothetical protein"  
/protein\_id="AAK9860.1"  
/db\_xref="GI:15458678"  
/translation="MTTKKANTITIEEYEMSEVDVNEAVNVEFTSDYCOLANSIYOSL  
KRFKFKMGSDGLIFITKSPSLVEGDIIGRDSOVNLRVIGIPENYLTNKSINL  
MNRKNCSDPFHDFDIYLDHVAAYADQKYNKIKEYPLKRALIHENMLYRLSEN  
EDDTLEKNTLYLTIMQVSEKTEPSEMDFMENKISEKIIFERGSMLNLDKSNYK"  
complement(752..1465)  
/gene="spri057"  
complement(752..1465)  
/gene="spri057"  
/codon\_start=1  
/transl\_table=11  
/label=spri057  
/product="Conserved hypothetical protein"  
/protein\_id="AAK9861.1"  
/db\_xref="GI:15458679"  
/translation="MFYKELLFDLIDHTLLDPDAEDVALYQLLKEGVADIAQKDY  
VPMKALKMDLELKISKQELVNTFRFSRFSRFGQEKDSFLAORTQFYLAQGQGLS  
GAHLDLSLEIRVDYLAATGTAIGTGRLAQGLVPYNQVFTISQLTQKRPDLF  
YEKIQQITAGSKSEKTIIMIGSLTDIAGQGNNAQIDTIWNPNNHLENHTDAQPTVEY



[illegible]

gene	complement(8122. .9855) /gene="ptsi" /note="spri062" complement(8122. .9855) /gene="ptsi" /EC_number="2.7.3.9" /codon_start=1 /transl_table=11 /label=spri062 /product="PTS enzyme I" /protein_id="AAK99866.1" /db_xref="GI:15458684" /translation="MTEMLKGIAASDGVAAKAVILYOPDLSPETIVEDTNAEARL DALAQSODELSVIREKAVGTGCGEAAQVDAHLMTLADDEMSIQIKETIRAKYNAE AGLEAVTDFTTIEGHEDNPPYMERADRIDVTKRYLANILGKLPNPSIDEVYI IAHLDTPSDTRAQDLDRKNFKAFTVNIIGRTSHSAIMATLILAAVLTGNNTLETYKDD ILANGITGEVITINPTDEOAAEFKAAGEAAKOKAEWALLKDQVYADSKHPELAN IGTRDEGVGNNVIAAEVAGLYRTPEFLYMDODPTEDEOXYEAKVAYEGNNGPVYVR TMDIGDKELPEFPMHEMNPFTGFRALRISISTGQPMFRTQIRALLRSLVSGOLRI MFPVALLKEPRAKAVFDEDEKANLLAEGVAVADNIQVIMETIPAAALADQAKVY DFTSITGNDLIQYMAADRMEQVSYLYQNPSTLRLINNVIKAAIEKRWGMGCEE MAGQQAQVPLLYVMGIDEFSSKATSVLRTSLMKKIDTAKMEEVANRALTECSTMEV LELQREYVND"	
gene	complement(9861. .10124) /gene="ptsh" /note="spri063" complement(9861. .10124) /gene="ptsh" /codon_start=1 /transl_table=11 /label=spri063 /product="histidine-containing phosphocarrier protein of the PTS"	
CDS	/protein_id="AAK99867.1" /db_xref="GI:15458685" /translation="MASKDFHYVAETGIIHARPATLLIQTASKFASDITLEYKKKSVNL KSIWGMSLVGCGADYTIASGEGADADALAIASEIMEKGCLA"	
BASE COUNT	2851 a 2091 c 1862 g 3516 t	
ORIGIN		
Query Match	99.0%: Score 2365.6; DB 1; Length 10320;	
Best Local Similarity	99.4%: Pred. No. 0;	
Matches 2374; Conservative	0; Mismatches 15; Indels 0; Gaps 0;	
1	TTCTTACGAGTTGGACGTGTATCAAGCTAGAACGGTTAAGCAAAATATTCGTTTCCCTA	60
Db		
7808	TTCTTACGAGTTGGACGTGTATCAAGCTAGAACGGTTAAGCAAAATATTCGTTTCCCTA	7749
61	TATAGATGGAATAACAGCGACGCCAATAACGAGAAATTTGACTCCTGATGAGGTTAGCAA	120
Db		
7748	TATAGATGGAATAACAGCGACGCCAATAACGAGAAATTTGACTCCTGATGAGGTTAGCAA	7689
121	GGGTGAAGAAATCATGCTGAGACCAATGCGTATCAAGATATACACACCAAGGCTATGTAC	180
Db		
7688	GGGTGAAGAAATCATGCTGAGACCAATGCGTATCAAGATATACACACCAAGGCTATGTAC	7629
181	TTACATGGCGACCACTATCATATTATTACAATGTTAGGTTCTCTTATGACGCTATCATCAG	240
Db		
7628	TTACATGGCGACCACTATCATATTATTACAATGTTAGGTTCTCTTATGATGCGATTTTCAG	7559
241	TGAAGATTACTCATGAAGAATCCAAACTATTAAGCTAAAGATGAGATATTTGTTAATGA	300
Db		
7568	TGAAGATTACTCATGAAGAATCCAAACTATTAAGCTAAAGATGAGATATTTGTTAATGA	7509

QY 301 GGTCAAGGCTGATATGTTATCAAGGTAGATGGAATAACTATGTTTACCTTAAGCATGC 360  
|||||  
Db 7508 GGTCAAGGCTGATATGTTATCAAGGTAGATGGAATAACTATGTTTACCTTAAGCATGC 7449  
QY 361 TGGCCACGGGGAATACGTCCTCAAAAGAGGAATCATCGACAAACAAAGCATGAG 420  
|||||  
Db 7448 TGGCCACGGGGAATACGTCCTCAAAAGAGGAATCATCGACAAACAAAGCATGAG 7389  
QY 421 TCACATCTGTAAGGTGGAACCTCAAGAAACGATGCTGCTGCTTGGACGCTTCGA 480  
|||||  
Db 7388 TCACATCTGTAAGGTGGAACCTCAAGAAACGATGCTGCTGCTTGGACGCTTCGA 7329  
QY 481 AGGACGCTTACTACAGATGATGTTATATCTTTAATGCTTGTGATATCATAGAGATAC 540  
|||||  
Db 7328 AGGACGCTTACTACAGATGATGTTATATCTTTAATGCTTGTGATATCATAGAGATAC 7269  
QY 541 TGGTATGCTTATATGCTTCCATGAGATCATTTACCATTTACATTCCTTAAGAAATGAT 600  
|||||  
Db 7268 TGGTATGCTTATATGCTTCCATGAGATCATTTACCATTTACATTCCTTAAGAAATGAT 7209  
QY 601 ATCAGCTAGCAGATTGGCTGCTCAGAAAGCCTTCTATCTGCTGAGAAATCTGTCAA 660  
|||||  
Db 7208 ATCAGCTAGCAGATTGGCTGCTCAGAAAGCCTTCTATCTGCTGAGAAATCTGTCAA 7149  
QY 661 TTCAGAACCTTATCGCCGACAAATAAGCATTAACCTTCAAGAACAACTGGGTACCTTC 720  
|||||  
Db 7148 TTCAGAACCTTATCGCCGACAAATAAGCATTAACCTTCAAGAACAACTGGGTACCTTC 7089  
QY 721 TGTAAACCATCCAGAAACATAAATCTAACACAGCAACAAAGAAACACTAACGCTCA 780  
|||||  
Db 7088 TGTAAACCATCCAGAAACATAAATCTAACACAGCAACAAAGAAACACTAACGCTCA 7029  
QY 781 AGCAAGTCAAGTAATGATGATGATGCTCTTGAAGACGCTCTACAAATGCTTTGAG 840  
|||||  
Db 7028 AGCAAGTCAAGTAATGATGATGATGCTCTTGAAGACGCTCTACAAATGCTTTGAG 6969  
QY 841 TCACAGCATGTAAGATCTGATGCTTGTCTTGTATCCAGACAAATCAAGATGGAAC 900  
|||||  
Db 6968 TCACAGCATGTAAGATCTGATGCTTGTCTTGTATCCAGACCAATCAAGATGGAAC 6909  
QY 901 AGCTAGAGGTGGACGAGTGCACAGGAGATCATTCACCTTACCTTACTCAAT 960  
|||||  
Db 6908 AGCTAGAGGTGGACGAGTGCACAGGAGATCATTCACCTTACTCAAT 6849  
QY 961 GTCGTAAATGGAAGAACGATGCTGATTTATTCCTTCTGTTATCTTCAAACTTGG 1020  
|||||  
Db 6848 GTCGTAAATGGAAGAACGATGCTGATTTATTCCTTCTGTTATCTTCAAACTTGG 6789  
QY 1021 GGTACCAATTCAGAGCCAGAACACCAAGTCCACAAACGACTCCGGAACCTAGTCCAG 1080  
|||||  
Db 6788 GGTACCAATTCAGAGCCAGAACACCAAGTCCACAAACGACTCCGGAACCTAGTCCAG 6729  
QY 1081 CCGGCAACCTGACCAAAATCTTAAATAGACCTCAAAATCTTCTTGGTATAGTACGTG 1140  
|||||  
Db 6728 CCGGCAACCTGACCAAAATCTTAAATAGACCTCAAAATCTTCTTGGTATAGTACGTG 6669  
QY 1141 ACGAAAAAGTTGGGAGATATGATTCGAGAAAGGCACTCTCTTATGCTTTTGC 1200  
|||||  
Db 6668 ACGAAAAAGTTGGGAGATATGATTCGAGAAAGGCACTCTCTTATGCTTTTGC 6609  
QY 1201 GAAAGATTTCACATCTGAAACCTGTAAAAATCTTGAAGCAAGTTTCAAAACAGAG 1260  
|||||  
Db 6608 GAAAGATTTCACATCTGAAACCTGTAAAAATCTTGAAGCAAGTTTCAAAACAGAG 6549  
QY 1261 TGTATACACACTTAACTGCTTAAAGAAAGAAATGTTGCTCCGAGACAAAGATTTTA 1320  
|||||  
Db 6548 TGTATACACACTTAACTGCTTAAAGAAAGAAATGTTGCTCCGAGACAAAGATTTTA 6489  
QY 1321 TGAATAAGCATATATCTGTTAACTGAGGCTCATAAAGCCTTGTGNAATTAAGGCTG 1380  
|||||  
Db 6488 TGAATAAGCATATATCTGTTAACTGAGGCTCATAAAGCCTTGTGNAATTAAGGCTG 6429  
QY 1381 TAAATTCGATTTCCAAAGCCTTAGACAAATTTATTAGAAGCCTTGATGATGATCGACTAA 1440

Db 6428 TAAATTCGATTTCCAAAGCCTTAGACAAATTTATTAGAACCTTGAAGATGAATCGACTAA 6369  
QY 1441 TAAAGAAAATGGTATGATTTATTTGGCATTCCTAGACCAATTAACCATCCAGAGCG 1500  
Db 6368 TAAAGAAAATGGTATGATTTATTTGGCATTCCTAGACCAATTAACCATCCAGAGCG 6309  
QY 1501 ACTTGGCAACCAAAATTCCTCAAAATTTAGTATCTGAAAGAGAAATCTGTATTTGCTCAAT 1560  
|||||  
Db 6308 ACTTGGCAACCAAAATTCCTCAAAATTTAGTATCTGAAAGAGAAATCTGTATTTGCTCAAT 6249  
QY 1561 AGCTGATTAAGTATACAGCTCAGATGCTTACATTTTGTATGAACATGATTAATCAGTGA 1620  
|||||  
Db 6248 AGCTGATTAAGTATACAGCTCAGATGCTTACATTTTGTATGAACATGATTAATCAGTGA 6189  
QY 1621 TGAAGAGATGCAATATGTAACGCTCATATGAGGCTTACTGCTGATGAGAAAGATG 1680  
|||||  
Db 6188 TGAAGAGATGCAATATGTAACGCTCATATGAGGCTTACTGCTGATGAGAAAGATG 6129  
QY 1681 CCTTCTGATTAAGAAAAGTTGACAGCTCAGCTTACTTAAAGAAAAGTATCCTTACC 1740  
|||||  
Db 6128 CCTTCTGATTAAGAAAAGTTGACAGCTCAGCTTACTTAAAGAAAAGTATCCTTACC 6069  
QY 1741 TCCATCTCCAGACGAGATGTTAAAGCAATCCAACTGAGATAGTGCAGACTATTTA 1800  
|||||  
Db 6068 TCCATCTCCAGACGAGATGTTAAAGCAATCCAACTGAGATAGTGCAGACTATTTA 6009  
QY 1801 CAATGCTGTAAGAGGAGAAAAGCAATTCACGCTGTCAGTCCATATATGTTGAGCA 1860  
|||||  
Db 6008 CAATGCTGTAAGAGGAGAAAAGCAATTCACGCTGTCAGTCCATATATGTTGAGCA 5949  
QY 1861 TACAGTTGAGGTTAAAGAGTGAATTTGATTTATTCCTCAAGATCATTAACATTAAT 1920  
|||||  
Db 5948 TACAGTTGAGGTTAAAGAGTGAATTTGATTTATTCCTCAAGATCATTAACATTAAT 5889  
QY 1921 TAAATTTGCTTGGTATGATGATCAGCATACAAAGTCCAAATGGCTATACCTTGGAGA 1980  
|||||  
Db 5888 TAAATTTGCTTGGTATGATGATCAGCATACAAAGTCCAAATGGCTATACCTTGGAGA 5829  
QY 1981 TTTGTTTGGAGATTAAGTACTAGTATAGAACACCTGAGCAAGCTCCCATTTCTAATGA 2040  
|||||  
Db 5828 TTTGTTTGGAGATTAAGTACTAGTATAGAACACCTGAGCAAGCTCCCATTTCTAATGA 5769  
QY 2041 TGGATGAGGCAATGCCAGTAGACATGTTTAAAGCAAGAAACACAGTGAAGATCCAAA 2100  
|||||  
Db 5768 TGGATGAGGCAATGCCAGTAGACATGTTTAAAGCAAGAAACACAGTGAAGATCCAAA 5709  
QY 2101 TAAAGACTTCAAGCGGATGGAAGAGCCAGTATAGAGAAACACTGCTGAGCCAGAACTCC 2160  
|||||  
Db 5708 TAAAGACTTCAAGCGGATGGAAGAGCCAGTATAGAGAAACACTGCTGAGCCAGAACTCC 5649  
QY 2161 TCAAGTAGAGACTGAAAGAAAGTGAAGAGCCCACTCAAGAAAGAGAAATTTGCTTGGCAA 2220  
|||||  
Db 5648 TCAAGTAGAGACTGAAAGAAAGTGAAGAGCCCACTCAAGAAAGAGAAATTTGCTTGGCAA 5589  
QY 2221 AGTAAGGATTTAGTGTGAAAGCCAAATGCAACAGAACTAGCTGTTTACGAATTA 2280  
|||||  
Db 5588 AGTAAGGATTTAGTGTGAAAGCCAAATGCAACAGAACTAGCTGTTTACGAATTA 5529  
QY 2281 TTTGACTCTTCAAAATTTATGATTAACATATGATCATGGAGAGCAAGAAATTTACTTGC 2340  
|||||  
Db 5528 TTTGACTCTTCAAAATTTATGATTAACATATGATCATGGAGAGCAAGAAATTTACTTGC 5469  
QY 2341 GTTGTAAAGAGAGTAATCTTCACTCTAGTATAGAGAAATTAAC 2389  
|||||  
Db 5468 GTTGTAAAGAGAGTAATCTTCACTCTAGTATAGAGAAATTAAC 5420

RESULT 6  
SPNEU1915 20035 bp DNA linear HTG 11-Jul-2001  
LOCUS SPNEU1915  
DEFINITION Streptococcus pneumoniae clone G54, \*\*\* SEQUENCING IN PROGRESS \*\*\*  
In Ordered pieces.

ACCESSION AL449937  
VERSION GI:11545162  
KEYWORDS HTG; HTGS\_PHASE2.  
SOURCE Streptococcus pneumoniae.  
ORGANISM Streptococcus pneumoniae  
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
Streptococcus.  
REFERENCE 1 (bases 1 to 20035)  
AUTHORS Dopazo, J., Mendoza, A., Herrero, J., Caldera, F., Humbert, Y.,  
Friedli, L., Guerrier, M., Grand-Schenk, E., Gaudin, C., de  
Francesco, M., Polissi, A., Buell, G., Feger, G., Garcia, E., Peitsch, M.  
and Garcia-Bustos, J.F.  
Annotated draft genomic sequence from a Streptococcus pneumoniae  
type 19F clinical isolate  
JOURNAL Microb. Drug Resist. 7 (2), 99-125 (2001)  
MEDLINE 21335329  
PUBMED 11442348  
REFERENCE 2 (bases 1 to 20035)  
AUTHORS Dopazo, J., Mendoza, A., Herrero, J., Caldera, F., Polissi, A.,  
Humbert, Y., Friedli, L., Guerrier, M., Grand-Schenk, E., Gaudin, C., de  
Francesco, M., Buell, G., Feger, G., Garcia, E., Peitsch, M. and  
Garcia-Bustos, J.F.  
Direct Submission  
JOURNAL Submitted (31-OCT-2000) Research Department, Glaxo Wellcome, S.A.,  
Severo Ochoa 2, 28760 Tres Cantos, SPAIN  
COMMENT \* NOTE: This is a 'working draft' sequence.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
FEATURES  
source 1. 20035  
/organism="Streptococcus pneumoniae"  
/serotype="19F"  
/db\_xref="taxon:1313"  
/clone="G54"  
BASE COUNT 6369 a 3635 c 4108 g 5909 t 14 others  
ORIGIN  
Query Match 97.5% Score 2328.2; DB 2; Length 20035;  
Best Local Similarity 98.5% Pred. No. 0;  
Matches 2353; Conservative 6; Mismatches 30; Indels 1; Gaps 1;  
QY 1 TTCTTACGAGTTGGAGCTGTATCAAGCTAGAACGGTTAAGGAAATATATCGTTTCTTA 60  
DB 17070 TTCTTACGAGTTGGAGCTGTATCAAGCTAGAACGGTTAAGGAAATATATCGTTTCTTA 17129  
QY 61 TATATGATGGAACAAACACGACGCAAAACGAGAGATTTGACTCTGATGAGTTACCAA 120  
DB 17130 TATATGATGGAACAAACACGACGCAAAACGAGAGATTTGACTCTGATGAGTTACCAA 17189  
QY 121 GCGTGAAGAGATCAATGCTGAGCAATCGTCATCAAGATTAACGACCAAGGCTATGTCCAC 180  
DB 17190 GCGTGAAGAGATCAATGCTGAGCAATCGTCATCAAGATTAACGACCAAGGCTATGTCCAC 17249  
QY 181 TTACATGGGACCACTATCATTTATTAACAATGTAAGTTCTTATGACGCTATCATCAG 240  
DB 17250 TTACATGGGACCACTATCATTTATTAACAATGTAAGTTCTTATGACGCTATCATCAG 17309  
QY 241 TGAAGATTTACTCATGAAAGATCAAACTATAGCTTAAAGATAGAGATATTTGTAATGA 300  
DB 17310 TGAAGATTTACTCATGAAAGATCAAACTATAGCTTAAAGATAGAGATATTTGTAATGA 17369  
QY 301 GGTCAGAGGTGATATGTATCAAGTATAGTGAAGTAACTATGTTTACCTTAAGGATGC 360  
DB 17370 GGTCAGAGGTGATATGTATCAAGTATAGTGAAGTAACTATGTTTACCTTAAGGATGC 17429  
QY 361 TGCCACGCGGATATAGTCGCTACAAAGAGAAATCAATCGACAAAACAGACATAG 420  
DB 17430 TGCCACGCGGATATAGTCGCTACAAAGAGAAATCAATCGACAAAACAGACATAG 17489  
QY 421 TCAACATCGTGAAGGTGAGTCTCCAAAGACATGATGCTGCTTGGCAGCTTGCGCA 480  
DB 17490 TCAACATCGTGAAGGTGAGTCTCCAAAGACATGATGCTGCTTGGCAGCTTGCGCA 17549

QY 481 AGACGGTATACATACAGATGATGTTATATCTTAATGCTTGTATATCATAGAGATAC 540  
DB 17550 GGGACGCTACACACACAGATGATGTTATATCTTAATGCTTGTATATCATAGAGATAC 17609  
QY 541 TGTGATGCTTATATATGCTTCTCCTAGAGATCATTTACATTTACTTAAGATGAGTT 600  
DB 17610 GGTGATGCTTATATATGCTTCTCCTAGAGATCATTTACATTTACTTAAGATGAGTT 17669  
QY 601 ATCAGCTAGCGAGTTGGCTGCTGAGAAAGCTTCTATCTGTGTGAGAGAAATCTGTCAA 660  
DB 17670 ATCAGCTAGCGAGTTGGCTGCTGAGAAAGCTTCTATCTGTGTGAGAGAAATCTGTCAA 17729  
QY 661 TTCAAGACCTATGCGGACAAAATAGCAATTAACATCTCAAGAACAAACGAGTACCTTC 720  
DB 17730 TTCAAGACCTATGCGGACAAAATAGCAATTAACATCTCAAGAACAAACGAGTACCTTC 17789  
QY 721 TGTATGCAATCCAGAACTACAAATTAACACACAGCAACACAACTAACACTCA 780  
DB 17790 TGTATGCAATCCAGAACTACAAATTAACACACAGCAACACAACTAACACTCA 17849  
QY 781 AGCAAGTCAAAAGTATATGATATGATAGTCTCTGTAAGACGCTTACAAACTGCTTGAG 840  
DB 17850 ARCAAGTCAAAAGTATATGATATGATAGTCTCTGTAAGACGCTTACAAACTGCTTGAG 17909  
QY 841 TCAACGACATGTATAGATCTGATGCGCTTGTGTGATCCAGCAAAATCCAACTGCAAC 900  
DB 17910 TCAACGACATGTATAGATCTGATGCGCTTGTGTGATCCAGCAAAATCCAACTGCAAC 17969  
QY 901 AGCTAGAGGTGCTTACAGTGCACACGAGATCATTTACACTTCACTTCTTACTCTCAAT 960  
DB 17970 AGCTAGAGGTGCTTACAGTGCACACGAGATCATTTACACTTCACTTCTTACTCTCAAT 18029  
QY 961 GTCTGAATTTGGAAGAACGAATCGCTGATATATTCCTCTGTTATGTTCAACCAATFG 1020  
DB 18030 GTCTGAATTTGGAAGAACGAATCGCTGATATATTCCTCTGTTATGTTCAACCAATFG 18089  
QY 1021 GGTACAGATTTCAAGGGCGACAGAACCAAGTCCCAACGCGCTCGGAACCTATGTCAGG 1080  
DB 18090 GGTACAGATTTCAAGGGCGACAGAACCAAGTCCCAACGCGCTCGGAACCTATGTCAGG 18149  
QY 1081 CCCGCAACCTGACCAATATCTTAAATATAGCTCAAAATCTTCTTGGTTAGTACGTGCT 1140  
DB 18150 CCCGCAACCTGACCAATATCTTAAATATAGCTCAAAATCTTCTTGGTTAGTACGTGCT 18209  
QY 1141 ACGAAAAGTTGGGGAAGATATGTTTCGAAAGAAAGGCAATCTCTGTTATGCTTTGC 1200  
DB 18210 ACGAAAAGTTGGGGAAGATATGTTTCGAAAGAAAGGCAATCTCTGTTATGCTTTGC 18269  
QY 1201 GAAAGATTTACCATCTGAACCTGTTAAATCTTGAAGAAAGCAAGTATCAAAACAGAGAG 1260  
DB 18270 GAAAGATTTACCATCTGAACCTGTTAAATCTTGAAGAAAGCAAGTATCAAAACAGAGAG 18329  
QY 1261 TGTTCACACACTTAACTGCTTAAAGAAAGAAATGTTGCTCTCTGACCAAGAAATTTTA 1320  
DB 18330 TGTTCACACACTTAACTGCTTAAAGAAAGAAATGTTGCTCTCTGACCAAGAAATTTTA 18389  
QY 1321 TGATTAAGCATATATCTGTTTAACTGAGGCTCATAAAGCCTTGTGTTTAAATTAAGGCTG 1380  
DB 18390 TGATTAAGCATATATCTGTTTAACTGAGGCTCATAAAGCCTTGTGTTTAAATTAAGGCTG 18449  
QY 1381 TAAATTCGATTTCCAGCCTTGAACAAATTAATTTGAAGCGCTGATGATGATGATGACCTTA 1439  
DB 18450 TAAATTCGATTTCCAGCCTTGAACAAATTAATTTGAAGCGCTGATGATGATGATGACCTTA 18509  
QY 1440 ATAAAGAAAAATTTGATAGATGATTTATTTGCAATTCCTAGCAACCAATTAACCATCAAGC 1499  
DB 18510 ATAAAGAAAAATTTGATAGATGATTTATTTGCAATTCCTAGCAACCAATTAACCATCAAGC 18569  
QY 1500 GACTTGGCAACCAAAATTTCAAAATTAAGTATATGATGAAAGCAAGTTCGATTTGCTCAAT 1559  
DB 18570 GACTTGGCAACCAAAATTTCAAAATTAAGTATATGATGAAAGCAAGTTCGATTTGCTCAAT 18629

QY	1560	TAGCGATTAAGATATCAACGTGAGATGGTTTACATTTTGTATGATAACATGATATATACAGTG	1619
Db	18630	TAGCGATTAAGATATCAACGTGAGATGGTTTACATTTTGTATGATAACATGATATATACAGTG	18689
QY	1620	ATGAAGGAGATGCATATGTAAACGCGCTCATANTGGGCCATGTGTCAGTGGATTTGGAAAAGATA	1679
Db	18690	ATGAAGGAGATGCATATGTAAACGCGCTCATANTGGGCCATGTGTCAGTGGATTTGGAAAAGATA	18749
QY	1680	GCCTTTCTGATTAAGGAAAAAGTTGCAAGCTCAAGCCATATCTAAGAAAAAGGTATCTTAC	1739
Db	18750	GCCTTTCTGATTAAGGAAAAAGTTGCAAGCTCAAGCCATATCTAAGAAAAAGGTATCTTAC	18809
QY	1740	CTCCATCTCCACAGCAGATGTTAAAGCAAAATCCAACTGGAGATGTGCGAGAGCATATT	1799
Db	18810	CTCCATCTCCACAGCAGATGTTAAAGCAAAATCCAACTGGAGATGTGCGAGAGCATATT	18869
QY	1800	ACAATCGTGTGAAGAGGGGAAAAACGAATTTCCACTGTTGCACTTCCATATATGTTGAGAC	1859
Db	18870	ACAATCGTGTGAAGAGGGGAAAAACGAATTTCCACTGTTGCACTTCCATATATGTTGAGAC	18929
QY	1860	ATACAGTTGAGGTTAAAAACGGTAATTGATTTATTCCTCATAGAGATCTATACATATA	1919
Db	18930	ATACAGTTGAGGTTAAAAACGGTAATTGATTTATTCCTCATAGAGATCTATACATATA	18989
QY	1920	TTAAATTTGCTGTGTTGATGTATCATACATATAAAGCTCCAAATGGCTATACCTTGGAG	1979
Db	18990	TTAAATTTGCTGTGTTGATGTATCATACATATAAAGCTCCAAATGGCTATACCTTGGAG	19049
QY	1980	ATTGTTTGGCAGATTAAAGTACTACTAGTGAACACCCCTGACGACGTCACATTTCTAATG	2039
Db	19050	ATTGTTTGGCAGATTAAAGTACTACTAGTGAACACCCCTGACGACGTCACATTTCTAATG	19109
QY	2040	ATGATTTGGGGCAATGCCAGTGCATGTGTTAGGCAAGAAAGACCAAGTGAAGATCCAA	2099
Db	19110	ATGATTTGGGGCAATGCCAGTGCATGTGTTAGGCAAGAAAGACCAAGTGAAGATCCAA	19169
QY	2100	ATAACAAGCTTCAACCGGATGGAAGGCCAGTGAAGAAAACACTGCTGACGACAGATCC	2159
Db	19170	ATAACAAGCTTCAACCGGATGGAAGGCCAGTGAAGAAAACACTGCTGACGACAGATCC	19229
QY	2160	CTCAAGTAGAGACTGAAAAAGTGAAGGCCCACTCAAAAGAAACAGAAAGTTTGTCTTGGA	2219
Db	19230	CTCAAGTAGAGACTGAAAAAGTGAAGGCCCACTCAAAAGAAACAGAAAGTTTGTCTTGGA	19289
QY	2220	AAATTAAGCGATTCTAGTCTGAAAGCCAAATGACAACAGAAACTCTAGCTGTTTACGAATA	2279
Db	19280	AAATTAAGCGATTCTAGTCTGAAAGCCAAATGACAACAGAAACTCTAGCTGTTTACGAATA	19349
QY	2280	ATTTCAGCTTCAAAATTTNTGATTAACAATAGATATCATGGCACAACAGAAAAATTAACCTG	2339
Db	19350	ATTTCAGCTTCAAAATTTNTGATTAACAATAGATATCATGGCACAACAGAAAAATTAACCTG	19409
QY	2340	CGTTGTTAAAGGAAGTAATCCTTCATCTGTAGTGAAGAAAAAATTAAC 2389	
Db	19410	CGTTGTTAAAGGAAGTAATCCTTCATCTGTAGTGAAGAAAAAATTAAC 19459	
RESULT 7			
AF340221			
LOCUS	AF340221	2535 bp	DNA linear BCT 22-MAY-2001
DEFINITION	Streptococcus pneumoniae Phpa (phpa) gene, complete cds.		
ACCESSION	AF340221		
VERSION	AF340221.1	GI:13447093	
KEYWORDS			
SOURCE			
ORGANISM	Streptococcus pneumoniae.		
	Streptococcus pneumoniae		
	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;		
	Streptococcus.		
REFERENCE	1 (bases 1 to 2535)		
AUTHORS	Zhang, Y., Masi, A.W., Barniak, V., Mountzouros, K., Hostetter, M.K. and Green, B.A.		
TITLE	Recombinant Phpa protein, a unique histidine motif-containing protein from Streptococcus pneumoniae, protects mice against		

FEATURES	source
JOURNAL	Intranasal pneumococcal challenge
MEDLINE	Infect. Immun. 69 (6), 3827-3836 (2001)
PUBMED	21246685
REFERENCE	11349048
AUTHORS	2 (bases 1 to 2535) Zhang, Y., Masi, A., Barniak, V., Mountzouras, K., Hostettler, M. and Green, B.
TITLE	Direct Submission
JOURNAL	Submitted (23-JAN-2001) Department of Bacteriology, Wyeth Lederle Vaccines, 21 Bailey Road, West Henrietta, NY 14586, USA
FEATURES	Location/Qualifiers
source	1..2535
gene	/organism="Streptococcus pneumoniae" /db_xref="taxon:1313"
CDS	1..2535 /gene="phpa" 1..2535 /gene="phpa" /note="histidine motif-containing protein" /codon_start=1 /trnas_table=11 /product="Phpa" /protein_id="AAK26629.1" /db_xref="GI:13447094"
BASE COUNT	880 a 518 c 533 g 604 t
ORIGIN	
Query Match	50.0%; Score 1434.4; DB 1; Length 2535;
Best Local Similarity	79.0%; Pred. No. 3.3e-296;
Matches 1724; Conservative	0; Mismatches 447; Indels 12; Gaps 1;
1	TTCTTCAGAGTTGGAGCTGATCAAGCTAGAGAAGGTTAAAGAAATATCGTGTCTTA 60
60	TTCTTCAGAGTTGGAGCTGATCAAGCTAGAGAAGGTTAAAGAAATATCGTGTCTTA 119
61	TATAGATGAAAAACAAGCGACGCAAAAAACGAGAAATTTGACTCCTGATGAGGTTAGCAA 120
120	TATAGATGAAAAACAAGCGACGCAAAAAACGAGAAATTTGACTCCTGATGAGGTTAGCAA 179
121	GGGTGAAGGAATCAATGCTGAGCAATTCGTATCAAGATTAACAGACCAAGGCTATGTCAC 180
180	GGGTGAAGGAATCAATGCTGAGCAATTCGTATCAAGATTAACAGACCAAGGCTATGTCAC 239
181	TTTCAGATGGGACCACTATCAATTAATTCAGAAAGTGAAGTTCCTATACAGCTATACATCAG 240
240	TTTCAGATGGGACCACTATCAATTAATTCAGAAAGTGAAGTTCCTATACAGCTATACATCAG 299
241	TGAAGAATTTACTCATGGAAGATCCAAACTATTAACCTAAAGATGAGATATTTGTTATGA 300
300	TGAAGAATTTGTTATGGAAGATCCAAACTATACACACTTAAGAGAGATATTTATCACTGA 359
301	GGTCAAGGTTGATATGTTATCAAGGTAGATGAAAAATTAATCTATGTTTACCTTAAGATGC 360
360	AATCAAGGTTGATATGTTATGTTATGTTAGTATGTAAGTAAATTAATCTTACCTTAAGATGC 419
361	TGCCACGCGGATTAACGTCCTACAAAAAGAGAAATCAATTCGACAAAAACAAGACATAG 420
420	AGCTCATCGGATATATGTCCTACAAAAAGAGAAATCAATTCGACAAAAACAAGACATAG 479

QY	421	TCACATCGTGAAGGTGGAACTCCAGAAAAGATGGTGGTCTGCTTGGCACGTTTCGA	480
Db	480	TCACATCGTGAAGGTGGAGATCCAGAAAAGATGGTGGTGGTCCCTTGGCACGTTTCGA	539
QY	481	AGGACGCTATCTCTCAGATGATGGTTATTCCTTAATAGCTTCGATCATATGAGGATAC	540
Db	540	GGGACGCTACACCAAGATGATGGTTATTCCTTAATAGCTTCGATCATATGAGGATAC	599
QY	541	TGGTGATCTTATATCGTTCCCTCAGAGATCATACATTTACATTCCTTAAGATGAGTT	600
Db	600	TGGTGATCTTATATCGTTCCCTCAGGCGACCATTTACATTTACATTCCTTAAGATGAGTT	659
QY	601	ATCAGCTGCGAGTGGGTGGTCAGAAAGCTTCATCTGGTGGCGAGAAATCTGTCAA	660
Db	660	ATCAGCTGCGAGTGGGTGGTCAGAAAGCTTCATCTGGTGGCGAGAAATCTGTCAA	719
QY	661	TTCAAGAACTTATGCGCGACAAAATAGCGATTAACCTTAAGAACTGGGTACCTTC	720
Db	720	TTCAAGAACTTATGCGCGACAAAATAGCGATTAACCTTAAGAACTGGGTACCTTC	779
QY	721	TGTAAAGCATTCAGAGCTACCAATACTTAACACAGCAACAGCAACACATTAACAGTCA	780
Db	780	TGTAAAGCATTCAGAGCTACCAATACTTAACACAGCAACAGCAACACATTAACAGTCA	839
QY	781	AGCAAGTCAAAAGTATGACATTTGATCTCTTTGAAAACAGCTCTACAACTCGTTTGAG	840
Db	840	AGCAAGTCAAAAGTATGACATTTGATCTCTTTGAAAACAGCTCTACAACTCGTTTGAG	899
QY	841	TCACAGACATGTACAATCTGATGGCCCTTGCTTTGATCCAGACAAATACAAAGTGAAC	900
Db	900	TCACAGCCATGTAAATCTGATGGCCCTTTATTTGACCCAGCCCAATACAAAGTGAAC	959
QY	901	AGCTPAGAGTGTTCAGTGGCCACAGGAGATCATTAACCTTATCCCTTACCTCACTCAAT	960
Db	960	CGCCAAATGGTGTCTGTACCGCAGCGAGACCAATTATCACTTATATCTTATTCACACT	1011
QY	961	GCTGGAATTTGGAAGAACGAATGCTGCTGATATTTCCCTTCGTTATCGTTCAACCATTG	1020
Db	1020	GTCACCTTTGGAAAATAATTTGGCTCGTATTTCCCTTCGTTATCGTTCAACCATTG	1079
QY	1021	GSTPCCAGATTCAGGGCAGAGCAACCACTCCACAAAGCATCTCCGAACTCATGTCAG	1080
Db	1080	GSTPCCAGATTCAGAGCCAGAGCAACCACTCCACATGACATCCGGAACCTCATGTCAG	1139
QY	1081	CCCGCAACCTCGCACCAATCTTTAAATATGACTCA-----ATTCTCTTGTGT	1128
Db	1140	TCCGCAACCTCGCACCAATCTCTCAACGAGCTCAAGGATTCGAATGATGAGAAATGGT	1199
QY	1129	TATGTCAGTGGTACGAAAAGTTGGGGAGAGATGATATTCGAGAAAGGCGATCTCTCG	1188
Db	1200	CAAGAAAGCTTTGGAAAAGTAAAGCGATGGTATATGCTTTGAGGAAATGAGGATTCCTCG	1259
QY	1189	TTATGCTCTTTGCGAAAATTTTACCATTTGAAATCTGTTAAAAATCTTGAAGAAAGTTATC	1248
Db	1280	TTATATCCCAAGCAAGATCTTTTCAGAGAAACAGCAGCAGCATTTGATAGCAAACTGGC	1311
QY	1249	AAAACAGAGAGTGTTCACACACTTTAACTGCTAAAAAAGAAATGTTGGCTCTCGTGA	1308
Db	1320	CAAGCAGGAAAAGTTTATCTATTAAGCTAGAGGCTTAAGAAAAGTACCTCCATCTAGTGA	1379
QY	1309	CCAGAAATTTATATTAAGCATATTAATCTGTTATACAGAGGCTCATTAAGAGCTTGTGTTG	1366
Db	1380	TAGAGAAATTTTCAATTAAGGCTTATGACTTACTAGCAGAGATTCATCAAGATTTACTTGA	1433
QY	1369	AAATTAAGGGTGTAAATTTCTGATTTCCAGGCTTAGACAAATTTATGAAGGCTTAAATGA	1428
Db	1440	TAAATTAAGGTGTGACAGAGTGTATTTTGAGGCTTTGGATTAACCTGTGGAACGACTCAAGA	1499
QY	1429	TGAATCGACTAATAAGAAAAATTTGGTAGATGATTTATTTGGCAATTCCTTAGCACAAATTAC	1488
Db	1500	TGTCTCAAGTATTAATCTCAAGTTAGTAGATGATATCTTGCCCTTTCTTGATTCCTGCAATTCG	1559

QY	1489	CCATCAGACGACCTTGGCAACCAATTTCTCAATTGATATACTGAAGCAAGTTGCG	1548
Db	1560	TCATCCGAGACCTTTAGGAAAACCAAAATGCGCAAAATTTACTCATCTGATGATGAAATTTCA	1619
QY	1549	TATTCGTCATTAAGCGTAATAGTATACAAACGTCACATAGGTATCACTTTTGTGATGAAACATGA	1608
Db	1620	AGTACGCCAAGTTGGAGGCGCAAGTATACAAACAGMAACGGTATATATCTTTGATCTCGTGA	1679
QY	1609	TATTAATCAGTGTGAGGAAGAGATGATATGTAACGGCTCATATGAGGCGCATATGACCTGGAT	1668
Db	1680	TATTAOCATGATGATGAGGGGAGTGCCTATGTAATCTCACATATAGACCATATGACCTGGAT	1739
QY	1669	TGGAAGACATAGCCCTTTCTGATAGGAAGAAAGTTGACAGCTCAAGCCTATACTAAGAAAA	1728
Db	1740	TAAAAAAGATATGTTGTCTCGAAGCTGAGAGACGGGACGCCAGGCTTATGCTAATGAAGAA	1799
QY	1729	AGATATCCACCTCCATCTCCAGACGCGAGATGTTAAACCAATCCAACTGGAGATATGTC	1788
Db	1800	AGGTTTGACCCCTCTTGACACAGACATCAGGATTCAGGAATACTGAGGCAAAAGGAGC	1859
QY	1789	AGCAGCTATTTTCAATCGTGTGAAGGGGAAAAAGCAATTCACACTGTTGACTCCATTA	1848
Db	1860	AGAACCTATCTCAACACCGGTAAAGACGATAGAAAGGTGCGCACTTGATCGTATCCCTTA	1919
QY	1849	TATGTTGACATACAGTTGAGTTAAAAACGGTAATTTGATTATTTCCATTCATTAAGGATCA	1908
Db	1920	CAATCTTCATATTAATCTGTAGAGTCAAAAAGCGTAGTTTAATCATACCTCATATAGACA	1979
QY	1909	TTACCATATATATTAATTTGCTTGTTGGTTGATGATCACACATACAAAGCTCCAAATGCGTA	1968
Db	1980	TTACCATATACATCAAAATTTGAGTGGTTTGACMAAGCCCTTATATGAGGCAACCTTAAGGGTA	2039
QY	1969	TACCTTGGAGATTTGTTGGACGATTAAGTACTACGACGTAACGACCCGCGAGAGCTTC	2028
Db	2040	TAGCTTGGAGATCTTTTGGCGACGTCAAGTACTATGTGCGAACATCCAAACGAAAGCTCC	2099
QY	2029	ACATTTCTATGATGATGAGGGCAATGCGCAGTAGTACATGTGTTAGGCAAGAAAGACACAG	2088
Db	2100	GCATTCAGTATATGATGTTTGGTATAGCGTACGACCATGTTCAAAGAAACAAAAATAGTCA	2159
QY	2089	TGAAGATCCAAATYANGAACCTTCAAGCGGATGAGAGCCAGTAGAGGAACACCTGCTGA	2148
Db	2160	AGCTGATACCATCAACAGGAAAAACCAAAAGAGAGAAACCTCGACAGAAAAACCTGA	2219
QY	2149	GCCAGAGTCCCTCAAGTAGAGA	2177
Db	2220	GGAGAGAAACCTTCGAGAGAGAGA	2242

Query Match

57.58; Score 1374.2; DB 6; Length 2523;



QY 2126 ----- 2125  
Db 2220 GGAAGAAACCCCTCGAGAAGAAACCAAGGAGAAACCAAGTCTCCAAACAC 2279  
QY 2126 -----CGAGTAGAGAAACACCTGCTGACGACGAAAGTCCCTCAAGTAGACTGAAA 2178  
Db 2280 AGAGGAACAGAGAAATCCAGAGATTCAGAAACCTCAGGTCGAGCTGAAAA 2339  
QY 2179 AGTAGAAGCCCACTCAAGAGAGAGAGTTCCTGCTGCGAAATTAACGAGTCTAGCT 2238  
Db 2340 GGTTAGAGAAAACTGAGAGGCTGAAGATTACTTGGAATAATCCAGATCAATTAT 2399  
QY 2239 GAAAGCCATGCAAGAACTAGCTGCTGTTAGAAATATTGACTCTCAATTAT 2298  
Db 2400 CAAGTCCATGCCAAGAGAGAGTCTACAGAGATTAATAATATTACTATTGGCACC 2459  
QY 2299 GGATTAACATAGTATCATGCGAGAGAGAAATTAATTCTGCTGTTAAAAAGAGTAA 2358  
Db 2460 GGAACAAATTAATATGATGCGAGAGCTGAAAACTATTGCTTATTAAGAGAGTAA 2519

RESULT 9  
AX343073 2647 bp DNA linear PAT 12-JAN-2002  
LOCUS AX343073  
DEFINITION Sequence 4 from Patent WO0198334.  
ACCESSION AX343073  
VERSION AX343073.1 GI:18152271  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.

REFERENCE  
1 Hamel,J., Ouellet,C., Charland,N., Martin,D. and Brodeur,B.  
AUTHORS Streptococcus antigens  
TITLE Patent: WO 0198334-A 4 27-DEC-2001;  
JOURNAL SHIRE BIOCHEM INC. (CA)  
FEATURES  
source 1. 2647  
Location/Qualifiers  
/organism="unidentified"  
/db\_xref="taxon:32644"  
/note="BVH-11"

BASE COUNT 934 a 538 c 556 g 619 t  
ORIGIN

Query Match 57.5%; Score 1374.2; DB 6; Length 2647;  
Best Local Similarity 73.9%; Pred. No. 2.4e-283;  
Matches 1819; Conservative 0; Mismatches 539; Indels 102; Gaps 2;

QY 1 TTCTTACGAGTTGGACTGTATCAAGCTAGAACGTTAAGGAAATATCGTTTCTTA 60  
Db 104 TCGTTATGAACTAGTTCGCTCAAGCTCAAACTGTAAAGAAATATCGTTTCTTA 163  
QY 61 TATAGATGAAACACGACGCAAAAAACGAGAAATTTGACTCTGATGAGTTAGCAA 120  
Db 164 TATAGATGAAACACGACGCAAAAAACGAGAAATTTGACTCTGATGAGTTAGCAA 223  
QY 121 GCGGAGAGATCAATGCTGCAAAATCGTCATCAAGATTAACCAACCAAGCTATGTCAC 180  
Db 224 GCGGAGAGATCAACCCGCAAAATCGTCATCAAGATTAACCAAGCTATGTCAC 283  
QY 181 TTCACATGCGGACCATATCTATTATCAATGTAAGTTCTTATGACGCTATCATCAG 240  
Db 284 CTCGATGAGAGACCATATCTATTATCAATGTAAGTTCTTATGACGCTATCATCAG 343  
QY 241 TGAAGATTAATCTATGAAGATCCAACTATAGCTAAAGAGATGATATTGTTAATGA 300  
Db 344 TGAAGATTAATCTATGAAGATCCAACTATAGCTAAAGAGATGATATTGTTAATGA 403  
QY 301 GGTCAGAGGATGATGTTATCAAGTAGAGTAAGTAAGTAAGTAAGTAAGTAAGTAAG 360  
Db 404 AATCAAGGATGATGTTATGATCAATTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAG 463

QY 361 TGCCACGCGGATTAACGTCCTACAAAAAGAGAAATCAATCGACAAAAACAAGACATAG 420  
Db 464 AGCTCATGCGGATTAATGTCGTTACAAAAAGAGAAATCAATCGGACAAAAACAAGACATAG 523  
QY 421 TCAACATGCTGAGGTGAGTCAACAGAAACGATGTCGTTGCTTGGCAGCTTGCA 480  
Db 524 TCAGCATGCTGAGAGAGGACTTACGAAACGATGTCGTTGCTTGGCAGCTTGCA 583  
QY 481 AGGAGCTATATCAATGATGATGTTATATCTTAATGCTTCTATATCAATAGAGATAC 540  
Db 584 GGGAGCTACACACAGATGATGTTATATCTTAATGATCTATATCAATAGAGATAC 643  
QY 541 TGTGATGCTTATATCGTTCCTCATGAGATCAATACATTAATCTTAAGATGAT 600  
Db 644 GGGAGATGCTTATATCGTTCCTCATGAGATCAATACATTAATCTTAAGATGAT 703  
QY 601 ATCAGCTAGCGAGTTGGCTGTCAGAAAGCTTCTATCTGTCGAGGAAATGTCAAA 660  
Db 704 ATCAGCTAGCGAGTTGGCTGTCAGAAAGCTTCTATCTGTCGAGGAAATGTCAAA 763  
QY 661 TTCAAGAACTATGCGCCGACAAATAGGATTAACACTTCAAGAACTGGATCCTC 720  
Db 764 TTCAAGAACTATGCGCCGACAAATAGGATTAACACTTCAAGAACTGGATCCTC 823  
QY 721 TGTAGCATCAGAGAACTAACAATACTACACAGCAACAGCAACACTAAGTCA 780  
Db 824 TGTAGCATCAGAGAACTAACAATACTAACAACAGCAACAGCAACACTAAGTCA 883  
QY 781 AGCAAGTCAAGATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 840  
Db 884 AGCAAGTCAAGATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 943  
QY 841 TCAACGATGATGAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 900  
Db 944 TCAACGATGATGAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1003  
QY 901 AGCTAGAGTGTGTCAGTGCACACGAGATCAATACACTTCACTTCTCTCAAT 960  
Db 1004 CGCCAGAGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1063  
QY 961 GTCTGAATGGAAGACGATGCTGATATATGCTGCTGCTGCTGCTGCTGCTGCTG 1020  
Db 1064 GTCTGAATGGAAGACGATGCTGATATATGCTGCTGCTGCTGCTGCTGCTGCTG 1123  
QY 1021 GGTACGATGATGAGGCGCAGACCAACCAAGCCGCAACCGGAACTAGTCCAG 1080  
Db 1124 GGTACGATGATGAGGCGCAGACCAACCAAGCCGCAACCGGAACTAGTCCAG 1183  
QY 1081 CCCGCAACCTGCAACCAATCTTAATAATAGACTCA-----ATTCTGTTGGT 1128  
Db 1184 TCGGCAACCTGCAACCAATCTTAATAATAGACTCA-----ATTCTGTTGGT 1243  
QY 1129 TAGTCAGCTGTTAGAAAGTTGGGAGAGATATGATTTGAGAAAGGAGATCTCG 1188  
Db 1244 CAAGAGAGCTGTTGAAAGTAGGCGATGATGATGATGATGATGATGATGATGATG 1303  
QY 1189 TTATGCTTTGCGAAAGATTACCATCTGAAACGTTTAAATCTTGAACCAAGTATC 1248  
Db 1304 TTATGCTTTGCGAAAGATTACCATCTGAAACGTTTAAATCTTGAACCAAGTATC 1363  
QY 1249 AAAACAAGAGAGTTTTCACACACTTAACTGCTAAGAAATAATGTTGCTCTGCA 1308  
Db 1364 CAAGCAAGAGAGTTTTCACACACTTAACTGCTAAGAAATAATGTTGCTCTGCA 1423  
QY 1309 CCAAGATTTTATGATTAAGCATATATCTGTTAATCTGAGGCTATTAAGGCTTGTG 1368  
Db 1424 TCGAGAAATTTTATGATTAAGCATATATCTGTTAATCTGAGGCTATTAAGGCTTGTG 1483  
QY 1369 AAATAGGCTGATATCTGTTTCCAAAGCTTGAACAAATATTAAGACCTGATGA 1428  
Db 1484 TAATTAAGGCTGATATCTGTTTCCAAAGCTTGAACAAATATTAAGACCTGATGA 1533  
QY 1429 TGAATCGACTAATTAAGAAATAATGATGATGATGATGATGATGATGATGATGAT 1488



[illegible]

ACCESSION	AF318954
VERSION	AF318954.1
KEYWORDS	GI:12744741
SOURCE	
ORGANISM	
REFERENCE	Streptococcus pneumoniae.
AUTHORS	Bacteria: Firmicutes; Lactobacillales; Streptococcaceae; Streptococcus.
TITLE	1 (bases 1 to 2457)
JOURNAL	Adamou, J. E., Heinrichs, J. H., Erwin, A. L., Walsh, W., Gayle, T.,
MEDLINE	Dornitzer, M., Dagan, R., Brewh, Y. A., Barron, P., Lathigra, R.,
PUBMED	21101045
REFERENCE	Identification and characterization of a novel family of
AUTHORS	pneumococcal proteins that are protective against sepsis
	Infect. Immun. 69 (2), 949-958 (2001)
	2 (bases 1 to 2457)
	1159990
	Adamou, J. E., Heinrichs, J. H., Erwin, A. L., Walsh, W., Dornitzer, M. and
TITLE	Johnson, S.
JOURNAL	Direct Submission
	Submitted (03-NOV-2000) Molecular Microbiology, Medimmune, Inc., 35
	West Watkins Mill Road, Gaithersburg, MD 20878, USA
FEATURES	Location/Qualifiers
source	1..2457
	/organism="Streptococcus pneumoniae"
	/serotype="4"
	/db_xref="taxon:1313"
gene	1..>2457
	/gene="phcB"
CDS	1..>2457
	/gene="phcB"
	/note="phcB; protective pneumococcal antigen"
	/codon_start=1
	/transl_table=11
	/product="pneumococcal histidine triad protein B precursor"
	/protein_id="AAK06759.1"
	/db_xref="GI:12744742"
	/translation="MKINKRYLAGSVAVIALSVCSYELGRYAGODKESNRVAYIDG DQAGKAENLTPEVSKREGINEAEOIYIKITDGYVSHGDHYHNGKVPYAISE ELMKPDNYOLKDSIDYNEIKGAVYVYLKDAHANIRKEIKRORER SHNNSRADNAVAARAQGRVTTDDGYIFASDIIETGDAYIPGDHYHYPKNEI SASLIAAEAVWNGKOSRPSSSSYANAPRLSENHNLVTPTPYHONGENISL LRELYAKPLSERHVESDGLIFDPAQITSTRAGVAVPHGNHIFIPYEQNSELEKRIA RIPIPLRASHNHVDSRPEPSPQPTPEPSPQAPASNPIDKLVKAEAVKAGDGV FEENGYSRYIPAKDLSAETAGIDSKLAKOESLSHKLGTKTDLPSSDRFYKAVDL LAR1HODLBNKGROVDFEALNDLNERKLVSSDKYLVEDILAFIAPIRHPRLGP NAOTYTRDDEIOVAKIAGKTAEGYIFEDRDTSDGDAVYVPHMTHSMIKKDIS EAEFAAQAATAXEKGLTPSTDHODSSNTAKGAETATYNNKAAKVPLDRMYNIQY TVEKNGSLIIPHYDHNIRFEWDEGLAEARGYLEDLATAVKYVEHPMERPHS DNGGNASDHVORNRKNGAQTNTQREKSEKPTQREKEETPREEKPOSEKPSPPPT EPEPSESEEPQVETEKVEEKREADELDGKIQDPIIKSNAKETLTGKNNMLTGT QDNNTTAAEAKEILLALLESK"
sig_peptide	1..87
	/gene="phcB"
misc_feature	247..264
	/gene="phcB"
	/note="Region: histidine triad"
misc_feature	616..633
	/gene="phcB"
	/note="Region: histidine triad"
misc_feature	718..792
	/gene="phcB"
	/note="Region: histidine triad"
misc_feature	940..957
	/note="Region: coiled-coil domain"
misc_feature	1336..1488
	/gene="phcB"
	/note="Region: histidine triad"
misc_feature	1657..1674
	/note="Region: coiled-coil domain"
misc_feature	1657..1674
	/gene="phcB"
	/note="Region: histidine triad"

```
misc_feature      1909..1926
                  /gene="phbB"
                  /note="Region: histidine triad"
repeat_region     525..714
                  /rpl_type=direct
repeat_region     1578..1758
                  /rpl_type=direct
BASE COUNT      832 a      496 c      533 g      593 t      3 others
ORIGIN
Query Match      41.5%; Score 991; DB 1; Length 2457;
Best Local Similarity 65.7%; Pred. No. 2e-201;
Matches 1609; Conservative 1; Mismatches 697; Indels 141; Gaps 6;

QY      1 TTCTTGAGATTGGAGCTGATCAAGCTAGACGGTTAAAGCAAA--TAATGGTGTTC 57
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      60 TTCTTGAGCTTGAGACCTTGACGTTACCAAGCTGTGACGATAGCAAAAGCTATAGAGTTGC 119

QY      58 CTATATAGATGAAAAACAAAGCGACGAAAAACGAGAAATTGACTCTGATGAGTTAG 117
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      120 TTATATAGATGATGATCAGGCTGTGCAAAAGCGAAAACTTGACACCAAGATGAGTCAG 179

QY      118 CAAGCGTGAAGGAATCAATCTGACCAATCGTCATCAATTAACAGACCAAGGCTATGT 177
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      180 TAAGAGGAGGAGGATCAACCCGACAAATTTGATCAAGATTACGATCAAGGTAATGT 239

QY      178 CACTTCACATGGCGACCACTATTATTAATGTAAGGTTCCTTATGACGCTATCAT 237
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      240 GACCTCTCATGGAGACCATTTATCTATTCTTAATGCGAAGGTTCTTATATGCCATCAT 299

QY      238 CAGTGAAGAAATTACTGTGAAGATCCAAACTTAAGCTTAAAGATAGAGATATTGTAA 297
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      300 CAGTGAAGAACTCCTCATGAAGATCCGAATTTATCAGTTGAAGGATTACAGCATTTGCAA 359

QY      298 TGAAGTCAGAGGTGATATCTTCAAGGTAGATGGAATTAATTAATTAATTAATTAAGA 357
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      360 TGAATCTAAGGTGATGATGATTAAGTAACCGTAATTAATTAATTAATTAATTAAGA 419

QY      358 TGTGCGCAAGCGGATTAACCTCCGTACAAAGAAGAAATCAATCGACAAACAAAGACA 417
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      420 TGTGCTCATAGCGGATTAATTTGCGACAAAGAAGATTAATTAACGTCAGAGCAAGACG 479

QY      418 TACTGCACATCGTGAAGGTGAACCTCCAAAGAAAGATGCTGTTCCTTGCGACAGTTC 477
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      480 CACTCATTAATCAT-----AAGTCAAGAGAGATTAATGCTGTGCTGACAGCAGAGC 530

QY      478 GCAGAGACGCTATACATAGATGATGTTATATCTTAAATGCTTGCATATCAATAGAGA 537
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      531 CCAAGGACGTTATACAAACGAGATGGGTATATCTTCAATGCAATCTGATATCAATTGAGA 590

QY      538 TACTGGATGCTTATATGCTTCCCTCATGAGATCATTAACATTACCTTAAGAAATGA 597
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      591 CACGGGTGATGCTTATATGCTTCCCTACAGCGACCATTTACATTACCTTAAGAAATGA 650

QY      598 GTTATACGTAAGCAGATTGGCTGCTGCAAGAGCTTCCATCTGCTGCGAAGAAATCTGTC 657
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      651 GTTATACGTAAGCAGATTGACTGCTGCGAAGAGCTTATGGAATG----- 695

QY      658 AAATTCAAGAACCTATACGCGCAAAAATAGCGATTAACACTTCAAGAACAAACTGGGATAC 717
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      696 -GAAGAGAGGATCTGCTCTTCTTCAAGTTCTAGTTTAATGCAAAATCCGCTCAACCAA 754

QY      718 TTCTGTAAGCAATCCAGGACCTCAATAATATAACACAAGAACACAAGCAACTATACAG 777
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      755 GATTGTGAGGAACCAATCT-----GACTGTCACTCCAACTTA 794

QY      778 TCAAGCAAGTCAAGATTAATGACATTGATGCTCTTGAAGACGCTCAACAACTGCTTT 837
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      795 TCAATCAAAATCAAGGGGAAAAAATTCAGACCTTTTACGTAATGTATGCTAAACCTT 854

QY      838 GAGTCAACGACATGTAGATCTGATGCGCTTGCTTGTGATCAGACCAAAATCACAGTGC 897
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      855 ATCAGAACGCCATGTGAATCTGATGCTTATTTTCGACCACAGCGCAAAATCACAAAGTCG 914
```

```
QY      898 AACAGTAGAGTGTTCAGTGGCCACAGAGATATTAACACTTCATCCCTTACTCTCA 957
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      915 AACCGCCAGAGGTGTAGCTCTCCCTCACTGTAACCATTTACACTTTATCCCTTATGACA 974

QY      958 AATGTCTGAATTTGGAAAGCAAGAAATCGCTGATTAATCCCTTGTATGTTCAAAACA 1017
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      975 AATGTCTGAATTTGGAAAGCAAGAAATGCTGATTAATCCCTTGTATGTTCAAAACA 1034

QY      1018 TTGGGTACCAAGATTTCAAGGCCAGAACACCAAGTCCCAACCGACCTCGAGACTAGTCC 1077
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      1035 TTGGGTACCAAGATTTCAAGGCCAGAACCAAGTCCCAACCGACCTCGAGACTAGTCC 1094

QY      1078 AGCGCCGCAACCTGACCAATATCTTAAATATAGACTCAATCTCTTGTGATGCTGCT 1137
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      1095 AAGTCCGCAACCGAGCTCCAAAGCAATCAATGATGGAA-----ATTGTCAAAGAACG 1148

QY      1138 GGTACGAAAAGTTGGGGAGAGATATGTTTGAAGAAAAGGGCATCTCTGTTATGTCTT 1197
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      1149 TGTTCGAAAAGTGAAGCGATGTTATGCTTTGAGAGAAATGAGATTCTCGTTATATCCC 1208

QY      1198 TCGGAAAAGATTTACATCTGAACCTGTTAAATCTTGAAGCAAGTTATCAAAACAGA 1257
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      1209 AGCCAAAGATCTTTCAGACAGAAACAGCAGCGCATTTGATAGAACTGGCCAAAGCA 1268

QY      1258 GAGTGTTCACACACTTAACTGCTTAAAAAAGAAAATGTTGCTCTGTCGACCAAGATT 1317
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      1269 AAGTTTATCTCATAGCTTAGAAGCAATTAAGAAACTGACCTCCCATCTGATGATGAGATT 1328

QY      1318 TTATGATAAAGCATATATATCTGTTAACTGAGGCTATAAAGCCTGTTGTTGAAATAAGG 1377
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      1329 TTACAAATAAAGCTTATGACTTACTAGCAAGAAATTAACCAAGATTTACTGATATAAAG 1388

QY      1378 TCGTAATTCGATTTCCAAACCTTAGCAAAATTAATTAAGACGCTTGATATGATGATGAC 1437
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      1389 TCGCAAGTTGATTTGAGGCTTTGATTAACCTGTTGGAACGACTCAAGGATGCTCAAG 1448

QY      1438 TAATTAAGAAAATTTGTTAATGATTTATGCAATTCCTGACCAATTAACCATCAGA 1497
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      1449 TGATTAAGTCAAGTATGATGATGATGATATTTCTGCTTCTTGAACCTGATGATGATGAC 1508

QY      1498 GCGACTTGGCAAAACCAATTTCTCAATTTAGATTAAGTAAGAGCAAGTTCGATTCCTCA 1557
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      1509 ACGTTTGAAGAAAACCAATTCGCAAAATTTACTACCTGATGATGATGATGATGATGAC 1568

QY      1558 ATTAGCTGATTAAGATTAACACGTCAGATGCTTACATTTTGAATGATGATGATGATGAC 1617
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      1569 GTTGGCAGGCAAGTACACAGCAGAGAGGTTATATCTTTGATGCTGCTGATATTAACAG 1628

QY      1618 TGATGAAGAGATGATATGATTAACGCTCATATGAGGCTATGATGATGATGATGATGATGAT 1677
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      1629 TGATGAGGGGAGATGCTATGATTAATCTGATGATGATGATGATGATGATGATGATGAT 1688

QY      1678 TAGCCTTTCTGATTAAGGAAAAGTTGCAAGCTCAAGCTATTAAGAAAAGGATATCCT 1737
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      1689 TAGTTTGTCTGAAGCTGATGAGAGGGGCAAGCCAGGCTTATGCTTAAGAGAAAGGTTGAC 1748

QY      1738 ACCTCCATCTCCAGACGCAATGTTAAAGCAAAATCCAATGAGATATGTCAGACAGTAT 1797
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      1749 CCGTCTCTTCAGACAGACCATGAGATTCAGGAAATTAAGGAGCAAAAGAGACAGAGATTAT 1808

QY      1798 TTACATACGCTGTGAAGGGGAAAAAGCAATTCACGCTGCTGCACTTCAATATATGTTGA 1857
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      1809 CTACAAACCGAGTGAAGACACAGTAAAGAGGTGCCACTTGATGATGATGATGATGATGAT 1868

QY      1858 GCATACAGTTGAGGTTAAACAGGTAATTTGATTAATCTCATTAAGGATCATTAACATAA 1917
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      1869 ATATACATGTAAGAGTCAAAAACGCTAGTTTAATCATACCTCATATATACCATACATAA 1928

QY      1918 TATTAATTTGCTTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1977
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      1929 CATCAAAATTTGAGAGTTTGAAGCAAGGCTTTATGAGGCACTTAAGGGGATATATCTTTGA 1988
```

QY	1978	AGATTGTTTGGCAGGATTAAGTACTAGCTAGCAACCCCTGCAGAACGCCATCTTAA	2037
Db	1989	GGATCTTTTGGCGACTGTCMASTACTAISTGCAACATCCAAACGACGCTCCCATTCAGA	2048
QY	2038	TGATGATGGGGCAATGCCAGTGCATGTGTTAGCGCAAGAAAGAACCCAGTGAAGATCC	2097
Db	2049	TAAATGTTTGGTAAACGCTAGCGCACCATGTTCAAGAAACAAAAATGTCACAGCTGATAC	2108
QY	2098	AAATTAAGAACTTCAAAAGGGATGA-----	2121
Db	2109	CAATCAAAACGAAAAACCAAGCGAGAGAAACCTAGACAGAAAAAAGCTGAGGAAGAAC	2168
QY	2122	-----AGACCAAGT	2130
Db	2169	CCCTGAGAGAAGAAACCGCAAAAGCAGAAACCAAGTCTCCAAAACCAACAGAGGAAGCC	2228
QY	2131	AGAGAAACACCTGCTGAGCCAGGAAGTCCCTCAAGTAAAGACTGAAAAAGTGAAGGCCA	2190
Db	2229	AGAAATAATCCACGAGGATCCAGAAAGAACCTTAGGTCGAGACTGAAAAGTTGGAAGAAA	2288
QY	2191	ACTCAAAAGACAGAGAAGTTTCTTGCAGAAAGTAACGAGTTCTAGCTGAAAGCCAAATGC	2250
Db	2289	ACTGAGAGAGAGCTGAAGATTACTTTGGAAAAATCCAGAGATCCAAATTATCAATCCAAATCC	2348
QY	2251	AACAGAAACCTAGCTGTTTACGAATAATTAATTGACTCTTCAAAATTATGATTAACAATAG	2310
Db	2349	CAAAAGACGCTCCACAGAGATTAAAAAAAATAATTAATTACTAATTTGGCACCCAGACAAACAATAC	2408
QY	2311	TATCTTGGCAGAGCAAGCAAAAAATTTCTGTCGCTGTGTTAAAGAAAGCTAA	2358
Db	2409	TATTATGGCAGAGCTGAAAAACATAATGGCGCTATTATTAAGGAGACTAA	2456

```

RESULT 11
AF318955          2517 bp      DNA       linear    BCT 11-FEB-2001
LOCUS             AF318955
DEFINITION        Streptococcus pneumoniae pneumococcal histidine triad protein D precursor (phtd) gene, partial cds.
ACCESSION         AF318955
VERSION           AF318955.1   GI:12744743
KEYWORDS
SOURCE            Streptococcus pneumoniae.
ORGANISM          Streptococcus pneumoniae.
REFERENCE         1 (bases 1 to 2517)
AUTHORS           Adamou,J.E., Heinrichs,J.H., Erwin,A.L., Walsh,W., Gayle,T., Dormitzer,M., Dagán,R., Brewal,Y.A., Barren,P., Lathigra,R., Langemann,S., Koenig,S. and Johnson,S.
                  Identification and characterization of a novel family of pneumococcal proteins that are protective against sepsis
                  Infect. Immun. 69 (2), 949-958 (2001)
TITLE             JOURNAL MEDLINE PUBMED REFERENCE AUTHORS
                  2 (bases 1 to 2517)
                  Adamou,J.E., Heinrichs,J.H., Erwin,A.L., Walsh,W., Dormitzer,M. and Johnson,S.
TITLE             Direct Submission
JOURNAL            Submitted (03-NOV-2000) Molecular Microbiology, MedImmune, Inc., 35 West Watkins Mill Road, Gaithersburg, MD 20878, USA
FEATURES
Source            location/qualifiers
                  1..2517
                   /organism="Streptococcus pneumoniae"
                   /serotype="4n"
                   /db_xref="taxon:1313"
                   1..>2517
                   /gene="phtd"
                   1..>2517
                   /gene="phtd"
                   /note="phtd: protective pneumococcal antigen"
                   /codon_start=1
                   /transl_table=11
                   /product="pneumococcal histidine triad protein D

```

	precursor"	
	/protein_id="AK06760.1"	
	/db_xref="GI:12744744"	
	/translation="MKINKRYLGSVALIALSVCSYELGRHQAQVKKESNRVSYIDG	
	DOAOAKENLTPEPEVSKREGINAEQVIKTTIDGQYVSHQDHYHNGKVPYDAIIE	
	ELMKDPNOYLKDSIDYNEIKRGVYIVDKQYVYLYLDAADANIRKKEIKROKH	
	SHNNGGSDOAVVAARQGRYTTDDCTIFMASQIIEEDGDATVPHGDHYHLPKNE	
	LSAEFLAAAEAYNNGKQGRSPSSSSSTINAPAPRSENNLTPTPHYHNGENISS	
	LRRLVPLPSERHVESDGLFDPAQITSTRARQVAPHONHFTIPEQMSLEKRI	
	ARILPLVRSNHWPDSRPSPQSPQSPQSPQSPQSPQSPQSPQSPQSPQSPQSPQ	
	KVGGVYEEENGVSRYIPAKDLSETAAGIDSKAKESLSHKLAGAKTDLPSPDRF	
	YKNGYDLARHLDLIPNKGROYDFELMDLLEKLVKQVSPDKLVLDIAPLAFIRH	
	PERGRGNAOITTYDDDEIOYAKLAGKTYTDEQYIFDPRDITDEGQAYVPHMTHSH	
	IKKQSLSEAEERRAAQATAKKEGLTPSTDPQSDNTFAKAEALYNNVKAAKVYDPR	
	MPYLYQITVEYKNGSLIIPHTDTHNKTFFEMFDEGLEAKGTLEDLQVYVYEH	
	PNEPNSHDNGEFGASDHVRKRNKYDDQSKPQEDKEHDEVSPTHESEDEKFNHAGLNS	
	ADNLNPKSTDEETEEDEAEEDTDEAEIPQVENSYNINKIADAEALLEKVPISIRONA	
	METUTGLKSSLLIGTRKDNMTISAEVDSLALLESQAPAIQ"	
	1. 87	
sig_peptide	/gene="phtd"	
	247. .264	
misc_feature	/gene="phtd"	
	/note="Region: histidine triad"	
	418. .495	
misc_feature	/gene="phtd"	
	/note="Region: coiled-coil domain"	
	619. .636	
misc_feature	/gene="phtd"	
	/note="Region: histidine triad"	
	1417. .1509	
misc_feature	/gene="phtd"	
	/note="Region: coiled-coil domain"	
	1678. .1685	
misc_feature	/gene="phtd"	
	/note="Region: histidine triad"	
	1930. .1947	
misc_feature	/gene="phtd"	
	/note="Region: histidine triad"	
	526. .717	
repeat_region	/rpt-type=direct	
	1597. .1779	
repeat_region	/rpt-type=direct	
	830 a 509 c 545 g 633 t	
BASE COUNT	830 a 509 c 545 g 633 t	
ORIGIN		
Query Match	41.48: Score 990; DB 1; Length 2517;	
Best Local Similarity	68.18: Pred. No. 3.2e-201;	
Matches 1478; Conservative	0; Mismatches 636; Indels 57; Gaps 5;	
OY	1 TTCTTACAGATTGGGACATGATCAAGCTATGAGAAAGGTTAAGGAAAA---TAACTCGTTC	57
Db	60 TTCTTATGAACCTTGGTCGTCACCAAGCTGGTCAGATTAGAAAAGATCTAATCGAGTTTC	119
OY	58 CTATATAGATGAAAAACAAGCGACGCAAAAAAGGAGATTGACTCCTGATGAGGTTAG	117
Db	120 TTAATATAGATGTTGATGTCAGGCTGGTCAAAAAGCAGAAAACCTTGACACCAAGTAAAGTCAG	179
OY	118 CAAGCGTGAAGGAATCAATGCTGTGAGCAAAATGCTATCAAGATTAAACAGACCAAGGCTATGT	177
Db	180 TTAAGAGGAGGAGGATCAACGCCCAACAAATGCTATCAAGATTACGATCAAGGTTATGT	239
OY	178 CACTTCACATGGCGACCACTATCATTTATTAACAATGTTAAGGTTCCCTATGACGCTATCAT	237
Db	240 GACCTCTCATGGAACCACTATTCATTTACTATAAAGGCAAGGCTCCCTTATGATGTCATCAT	299
OY	238 CAGTGAAGAATTTACTATGAAGAATCCAAACTATTAAGCTAAAGAATAGAGATGAGTTGTTAA	297
Db	300 CAGTGAAGAAGCTCCTCATGAAGAATCCGAATTTATCAAGTGAAGATTCAGACATTTGCAA	359
OY	298 TGAAGTCAAGGATGATATGTTATCAAGGTAGATGAAATAATATCATGTATTACCTTAAAGA	357

Db	360	TGAATCAAGG	GTGTTATATGTTTCAAGTAGTGGAAATCTATGTTTACCTTAAGGA	419
Oy	358	TGCTGCCACG	CGGATACGTCGTAACAAAGAGAAATCAATGCACAAAACAGAGCA	417
Db	420	TGCAGCTCAT	CCGATTAATATTCGACAAAGAGAGATTAACGTCAGGAAGCAGAGCA	479
Oy	418	TAGTCACATC	GTGAAAGTGGAACTCCAGAAACGATGGTGGCTGTGGCTTGCAAGCTTC	477
Db	480	CAGTCATAA	TACGCGGGGTGTTCT-----AACGATCAAGCAGTAGTTGCACCGAGGC	533
Oy	478	GCAAGGACGC	ATCTACATGATGTTATATCTTAATGCTGTCGATATCATATGAGGA	537
Db	534	CAAAGAGC	ATACACGATATGTTATATCTTCAATGATCATGATCATATTAAGGA	593
Oy	538	TACTGTGATG	CTTAATATTCGTTCCATATGAGATCATATACATATCAATCCTAAGATGA	597
Db	594	CACGGGTGAT	GTCTTATATGTTCTCTCAGCGCACATTAACATTAATTCCTTAAGATGA	653
Oy	598	GTTATACAT	AGCAGTTGGCTCTCGAAGCCTTCTATCTGTGAGAGAAATCTGTC	657
Db	654	GTTATACAT	AGCAGTTAGCTGCTGTCAGAGCT-----	688
Oy	658	AAATTCAGA	ACCTATGCCGACAAAATAGCGTAACTTTAAGAACAACTGGGTACC	717
Db	689	-----ATTG	AATGGAATGGAAAGGAGGATCTGCTCTTCAAGTTTACTTATATATGC	740
Oy	718	TTCTGTAA	GCATACAGACCTCAAAATCTCAACACACACAGACAGACACTAATAG	777
Db	741	AAATTCAC	AGCTACACAGATTTGTCAGAGAACCAATCTGATGTCACCTCAACTTA	800
Oy	778	TCAAGCA	AGTCAAAATGATGATGATGATGATGATGATGATGATGATGATGATGATG	837
Db	801	TCA---	AAATCAAGGGGAAACATTTCAAGCCTTTAGCTGAATGTATGCTAAACCTT	857
Oy	838	GAGTCAC	AGCATGTATGATATCTATGCTGTGCTTGTATCCAGACAAATCACAGTCG	897
Db	858	ATCAAGAA	CGCATGTGGAATCTATGCTCTTATTTTGACCCAGGCAAAATCACAAAGTCG	917
Oy	898	AACAGCTAG	AGGTTTGCATGTCACACAGGAGATCATATACACTCATACCTTACTCTCA	957
Db	918	AACGCCAG	AGGTGTACCTGTCCCTCATGTAAACCATTAACACTTATCCCTTATGAAACA	977
Oy	958	AATGCTGA	ATTTGGAGACGAATTCGCTGATTTATCCCTTGTCTTATGCTTAAACCA	1017
Db	978	AATGCTGA	ATTTGGAGAAAGCAATTCGCTGATTTATCCCTTGTCTTATGCTTAAACCA	1037
Oy	1018	TTGGGTAC	CAAGATTCAGGCCAACAACACAGTCCACACACGACTCCGGAACCTAGTCC	1077
Db	1038	TTGGGTAC	CAAGATTCAGGCCAACAACACAGTCCACACACGACTCCGGAACCTAGTCC	1097
Oy	1078	AGGCCCA	CACTGCACCAATCTTAAATATAGCTCAAAATCTTC-----TTT	1125
Db	1098	AAGTCCCA	CACTGCACCAATCTTAAATATAGCTCAAAATCTTC-----TTT	1157
Oy	1126	GTTAGTCA	CTGTGTAGGAAAGTGGGGAAGGATATGATTTGGAAGAAAGGGGACATTC	1185
Db	1158	GGTCAAA	GAAGCTGTGGAAAGTAGGCGGTATGTCTTTGAGGAATGAGAGTTTC	1217
Oy	1186	TGCTTAT	GTCTTTGCGAAGATTTACCATCTGAACCTGTTAAAAATCTTGAAGCACTT	1245
Db	1218	TGCTTAT	GTCTTTGCGAAGATTTACCATCTGAACCTGTTAAAAATCTTGAAGCACTT	1277
Oy	1246	ATCAAA	CAAGAGTGTTCACACCTTAACTGTAAAGAAAGAAATGTTGCTCCCG	1305
Db	1278	GGCCAA	GAGAGAGTTTATCTATAGCTAGGAGTAAAGAAATGCTCCATCTTG	1337
Oy	1306	TGACCA	GAATTTATGATTAAGCATATATCTGTTAACTGAGGCTCATTAAGCCTGTT	1365
Db	1338	TGATCGA	GAATTTTACATTAAGCTTATGACTTACTAGCAAAATTCACCAAGATTTCT	1397
Oy	1366	TGNAAT	TAAGGCTGTATATTTGATTTCCAGGCTTATGACAAATTTATTAAGACGCTTGA	1425

Dd	1398	TGATATAAAGTCGCACAACTGGATTGTTGAGCCTTGGATTAACCTGTGGGAACAAGCACTCAA	1457
Oy	1426	TGATGAATCGCATTAATAAAGAAAATTTGGTAGATGATTATTTAGCCATCTCAGCACCAAT	1485
Dd	1458	GGATGTGCCAAGTGATATAAGTCAACTAGTSGATATATTCCTGGCTCTTAAGCTCCGAT	1517
Oy	1486	TACCATTCCAGAGCAGCTTGGCAACCAATTCCTCAATTGAGTACTGAAGACGAAGT	1545
Dd	1518	TGCTATCTCCAGAACCTTTAGGMAAACCAATGCGCAAAATTAACCTCACTGATGATGAGAT	1577
Oy	1546	TGCTATTCGTCATTAAGTGGATAGATATACAAACGTCACATGGTTTCAATTTTTTGATGAGACA	1605
Dd	1578	TCAAGTACCGCAAGTTGGCAGCGCAAGTACACACAGAAACGGTATATCTTTGATCTTCG	1637
Oy	1606	TGATATTAATCAGTGTAGAAGAGATGCATATATGACGCTCATATGSGCCATATGCACTG	1665
Dd	1638	TGATATTAACCAAGTATGAGGGGAGATGCTATGTAACTCCACATATGACCATAGSCACTG	1697
Oy	1666	GATTGGAAAAAGATAGCCCTTCTGATTAAGSAAAAAGTTGCAAGCTCAAGCCTATACTTAAGA	1725
Dd	1698	GATTAAAAAAGATGATTGTCTGAAGCTGAGAGACGCGCAGCCAGGCTTATGCTTAAGA	1757
Oy	1726	AAAAGTATCCACTCCATCTCCAGACGAGATGTTAAAGCAATCCAAGTGGAGATAG	1785
Dd	1758	GAAAGTTTGAACCCCTCTTGACAGACCATTCAGGATTCAGAAATFACTGAGGCAAAAAG	1817
Oy	1786	TGCACAGCTATTTTCAATCGTGAAGGGGAAAAAGCAATTCACACTCGTTGCACTCC	1845
Dd	1818	AGCAGAACGTCATCTACAAACCGGTAAAGCAGCTTAGAAGGTGCCACTTGAATCGTATCC	1877
Oy	1846	ATATATGGTTGACATGAGTTGAGGTTAAAAACGGTAAATTTGATTATTCCTCATTAAGA	1905
Dd	1878	TTACATATCTTCAATTAAGTGTAGAAGTCAAAAAGCGTAGTTAATCAATCACTCATTAAGA	1937
Oy	1906	TCATTAACCATATATTAATTTGCTTGGTTGATGATACACATACAAAGCTCCAATAG	1965
Dd	1938	CCATTATACCATTAACATCAAAATTTGAGTGGTTTGACCAAGCCCTTAATGAGGCAACCTAAGG	1997
Oy	1966	CTATACCTTTGGAAGATTGTTGGCAGCATTAAGTACTACGTAGAACAACCTGAGCAAGC	2025
Dd	1998	GTAATCTCTTGAGATCTTTTGGCAGCTGTCAGATCATGTGCAACATCCAAAGCAAGC	2057
Oy	2026	TCCACATTCCTAATGATGAGATGGGGCAATGCCAGTGAAGCATGTGTTAGGCCAAGAAAGCCA	2085
Dd	2058	TCCGATTCAGATTAATGATGTTTGGTAAGCTAGACGACCATGTGCTAAATAATTAAGTAGA	2117
Oy	2086	CAGTAAGAATCCAAATAAGAACTTAAAGCGGATGAAGAAGCCAGTAGAGAAACACCTGC	2145
Dd	2118	CCAAACAGATTAACCTGATGAAGATGAAGAACATGATGAAGTAACTGAGCCACCACTACCC	2177
Oy	2146	TGAGCCCAAG 2156	
Dd	2178	TGAATCTGATG 2188	
RESULT 12			
AEO07403			
LOCUS		11931 bp	DNA linear BCT 31-AUG-2001
DEFINITION		Streptococcus pneumoniae TIGR4	section 86 of 194 of the complete genome.
ACCESSION		AEO07403	AE005672
VERSION		AEO07403.1	GI:14972469
KEYWORDS			.
SOURCE			Streptococcus pneumoniae TIGR4.
ORGANISM			Streptococcus pneumoniae TIGR4
REFERENCE			Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Streptococcus.
AUTHORS			1 (bases 1 to 11931) Tetzelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D., Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J., Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D., Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D., Holtzapple E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,

TITLE  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
2 (bases 1 to 11931)  
Tettelin,H., Nelson,K.E., Paulsen,I.T., Eisen,J.A., Read,T.D., Peterson,S., Heidelberg,J., DeBoy,R.T., Haft,D.H., Dodson,R.J., Durkin,A.S., Gwinn,M., Kolonay,J.F., Nelson,W.C., Peterson,J.D., Umayam,L.A., White,O., Lewis,M.R., Radune,D., Holtzapfle,E., Khouiri,H., Wolf,A.M., Uteerback,T.R., Hansen,C.L., McDonald,L.A., Feldblyum,T.V., Anginoli,S., Gesuwan,P., Hickey,E.K., Holt,I.E., Loftus,B.J., Ujval,M.L., Yang,F., Smith,H.O., Venter,J.C., Dougherty,B.A., Morrison,D.A., Hollingshead,S.K. and Fraser,C.M.  
Complete genome sequence of a virulent isolate of Streptococcus pneumoniae  
Science 293 (5529), 498-506 (2001)  
11463916  
2 (bases 1 to 11931)  
Tettelin,H., Nelson,K.E., Paulsen,I.T., Eisen,J.A., Read,T.D., Peterson,S., Heidelberg,J., DeBoy,R.T., Haft,D.H., Dodson,R.J., Durkin,A.S., Gwinn,M., Kolonay,J.F., Nelson,W.C., Peterson,J.D., Umayam,L.A., White,O., Lewis,M.R., Radune,D., Holtzapfle,E., Khouiri,H., Wolf,A.M., Uteerback,T.R., Hansen,C.L., McDonald,L.A., Feldblyum,T.V., Anginoli,S., Gesuwan,P., Hickey,E.K., Holt,I.E., Loftus,B.J., Ujval,M.L., Yang,F., Smith,H.O., Venter,J.C., Dougherty,B.A., Morrison,D.A., Hollingshead,S.K. and Fraser,C.M.  
Direct Submission  
Submitted (29-JUN-2001) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA  
Location/Qualifiers  
1. 11931  
/organism="Streptococcus pneumoniae TIGR4"  
/strain="TIGR4"  
/db\_xref="taxon:170187"  
129. .476  
/gene="SP0995"  
655. .795  
/gene="SP0996"  
/note="This region contains a match to at least one other gene that is not full length, and is not the result of a sequencing artifact; similar to GP:5019553; identified by sequence similarity; putative; IS630-Spn1, transposase Orf2, truncation"  
956. .1156  
/gene="SP0997"  
956. .1156  
/gene="SP0997"  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAK75114.1"  
/db\_xref="GI:14972470"  
/translation="MVASASSTSTQAEQVDSKELRALSQELDRKALATVSDPK IDAKKAVLDKAKPEDSALTE"  
1303. .1602  
/gene="SP0998"  
1303. .1602  
/gene="SP0998"  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAK75115.1"  
/db\_xref="GI:14972471"  
/translation="MTETNSVPAGYIVVSLALGVIAFWLIRRKKESEIOOLSTELL KVLGOLDKRAKVKLAKQNLDETLDFAVENSAETETKRLVEELKAIIDKIK"  
1833. .2540  
/gene="SP0999"  
1833. .2540  
/gene="SP0999"  
/note="similar to GP:6714962; identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="cytochrome c-type biogenesis protein CcdA"  
/protein\_id="AAK75116.1"  
/db\_xref="GI:14972472"  
/translation="METIVFLISVFLAGVLSFSPICRPLPYAGILLDDQESAKSF SLFGKVLMSGLIRLCTAGISLFTLGGAGFGHILTANMRIGWGAIIILGL HOMEIFHLKLEVOKSFTEFKRSDSNRYSAPLGLITTFSGWTPCIGPVLSSVALAAS

gene  
CDS  
2551. .3108  
/gene="SP1000"  
/codon\_start=1  
/transl\_table=11  
/product="chlorodoxin family protein"  
/protein\_id="AAK75117.1"  
/db\_xref="GI:14972473"  
/translation="KKVMEAGSLISLVLMACGEETKQTAQAPKQOTYQQA VGGAPDITQSMQKEVLSDFGKRYKLFMAWCGPCPKSPPELMELAKDRDF EILVIVAPGIGETVEQFPQWFOEQGKIDPIYDRTATFQYQIIRSIPTYLIDS OGKIKIGFAGISNADAARKEMN"  
3482. .4873  
/gene="SP1001"  
3482. .4873  
/gene="SP1001"  
/note="identified by match to PFAM protein family HM PF00324"  
/codon\_start=1  
/transl\_table=11  
/product="amino acid permease family protein"  
/protein\_id="AAK75118.1"  
/db\_xref="GI:14972474"  
/translation="MNIFRTKNVSLDKTEMRHKLMDLILIGICAMVGTVTGT AATLAGPALVTSIVSALCYGSLAPFAEASRPATGASVLYLALGEPFAMLVG WTMTEFWTATISGVAAGAAVEFKLLSGYIAPFOALNGTFNPAGPVDLPLIVLY LVSIVLINAKAALRPNISIVILKFSALALVGVINIKDKNSNRPYFGGITYGA STGIMAGSLMEFGFLGFEISMAVDEKTPQKNIPRIVISLIVTLAVLVLT GVVYSHLVNDVAFAFALRSVIGISMAVYSLVILITVICISMTYALSMYISLAR DGLVPAERKELTKSKIRKNATILTLGLASVAAGMPLASIAFLNICTLAVLIMAY GLIRLRKKGMPKAGEFETPLVPLIILISICSPMLQYNNMTWIAFLVLSGII VETGYKXSTIEE"  
5016. .5933  
/gene="SP1002"  
5016. .5933  
/gene="SP1002"  
/note="identified by match to PFAM protein family HM PF01297"  
/codon\_start=1  
/transl\_table=11  
/product="adhesion lipoprotein"  
/protein\_id="AAK75119.1"  
/db\_xref="GI:14972475"  
/translation="MKRONFLIVLISVFLCTGACGQKESOTGKMTVSPYIVAM VKESGDLNDVRAIQSSGHSFSPSANDIAITDADFTVHSHTLSEAGSDLPNIK KSKYVLEASGEKTLERYPGLEDVAGGVDEKTLDPHTWLDPEKAGEEQIADKL SEVDSEHKETVQKNAQAFIKKAQELTKFQPKFKAQKTEVTQHTAFSTYAKRFGIN OLGIAGISPEQEPSPROLTEIOEFVKYTKVITFTESNASSKVAETLVKSTVGKLT LNPESDPNDKTYLENLENNISLAELK"  
5941. .8460  
/gene="SP1003"  
5941. .8460  
/gene="SP1003"  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAK75120.1"  
/db\_xref="GI:14972476"  
/translation="MKINKRYLGSVAVLAVSCSYELGRHQAGOVKRESNRYSIDG DOAGOKALENLPDEVSKEGINABQIVIKITDQGYVSHGDIHYVYNGKVPYDIIE ELLMKDRNYQKSDIYNEIKGYIVYVDKYYVYLDAADANDIRKEIKRKOEH SHNKGGSNDQAVVAARQGRYTTDDGTFTASDIITEDTGAYTPHGDRHTPKNE LSASELAAEAERVHVKQSSRPSSSSNNANAPVPHGNHMYHFIPEQOSELEKRI LRLRLYKPLSERVESDGLIFDPAQITSRTARAVAPDHGNHMYHFIPEQOSELEKRI ARIILPKRSHHWVDSRBPQSPQPEPSPOVAPNPQAPSNPIDKLVKEAVR KVGQGVPEENSVARSYIPAKDLISATAGIDSKLAKOESLSHKIGAKKTDPSDRE YNKAYDLARIHQDLLDNKQVDFPEALDNLEKLYPSKRYLVLDIATFLPIKH PERIGKPAQOITTYDDDEQVAKLQAGKTTDEGYTFDPDITSDGDVAVTPHMTSHW IKKDSLSBAERAAQAQAKEKGLTPPSIDHDSGNTAKAKAEAIYNRYAKAKVPLDR







QY	838	GAGTCACGACATGTAAACATGTGATGGCTCTGTTTGGATTCACGACAAATTCACAAAGTC	897
Db	201147	ATCAGAACGCATGTGGAATCTGTATGGCTTTATTTTGGACCCAGCCCAATTCACAAAGTCG	201206
QY	898	AACAGCTAAGAGTGTTCGACGTGCACACGAGAGATCATTTACCTTCATCCCTTACTCTCA	957
Db	201207	AACCGCAGAGGTGTAGGTGTCCTCAATGGTAACCATTTACCACTTTATCCCTTATGAACA	201266
QY	958	AATGTCTGAATTTGGAAGACGAATTCGCTGTAATTAATCCCTTCGTTATTCGTTCAAAACA	1017
Db	201267	AATGTCTGAATTTGGAAGAACGAATTCGCTGTAATTAATCCCTTCGTTATTCGTTCAAAACA	201326
QY	1018	TTGGGTACCCAGATTCAGGGCCAGAACCAACCAAGTCCCAACGCACTCCGGAACCTAGTCC	1077
Db	201327	TTGGGTACCCAGATTCAGAACCCAGAACCAAGTCCCAACGCACTCCCAACCTAGTCC	201386
QY	1078	AGGCCCGCAACCTCCACCAATATCTTAAATAAGACTCAA-----ATTCTTCTTT	1125
Db	201387	AAGTTCGGAACCTCCACCAATATCTCTCAACCAAGCTCCCAAGCAATTCATATGAAATTT	201446
QY	1126	GCTTATGTCAGCTGTTACGAAAAGTTGGGGAAAGATATGTAATTCGAAGAAAAAGGCACTTC	1185
Db	201447	GGTCAAAAGAACCTGTTGCAAAAAGTAGGCGATGGTTATGCTTTGAGAGAAATGGAATTTTC	201506
QY	1186	TGCTTATGTCCTTTCGGAAGAATTACATCTGAAACGTGTTAAATTCCTTAAGACAAATT	1245
Db	201507	TGCTTATATCCCAACCAAGGATCTTTTCAGAGAAACAGACAGCGGCAATGTATGACAAACT	201566
QY	1246	ATCAAAACCAAGAGATGTTTTCACACACTTTAAGCTCTCAAAAAAGAAATGTGCTCTCTCG	1305
Db	201567	GGCCAGAGAGAAAGTTTATCTCTATAGCTATGAGAGCTAABAAAAATGACCTCCCATCTAG	201626
QY	1306	TGACCAAGAATTTTATGATAAGCATTAATCTGTTAAGTACTGAGGCTCATTAAGCCTTTGTT	1365
Db	201627	TGATCGAAGAAATTTTACATTAAGGCTTATGACTTACTAGCAAGAAATTCACCAAGTTTACT	201686
QY	1366	TGMAATTAAGGGGTGTATATCTGATTTCCAAAGCCTTAGCAAAATTTATTAACGCTTGAA	1425
Db	201687	TGATTAATTAAGGTGTGACAGCTTTGATTTTGAAGCTTTGGTAACCTGTTGGAAAGACTCAA	201746
QY	1426	TGATGAATCGCAATTAATAAGAAAAATGTGATGATTAATTTGGCATTCCTACACCAAT	1485
Db	201747	GGATGTCTCAGTATTAAGTCAAGTATAGTGAAGATATCTTGCTTCTTACTCTCCGAT	201806
QY	1486	TACCATTCACAGGAGCTTGGCAAACTTCATCAATTGAGTATTAAGTGAAGACGAAGT	1545
Db	201807	TGCTATCTCAGAACGTTTAGAAAAACCAATTCGCCAATTAATCTACACTGATGATGAGAT	201866
QY	1546	TGCTATTTGCTCAATTTAGCTGATTAAGTATACAAGCTCAGATGTTTACTATTTTGATGAACA	1605
Db	201867	TCAAGTATCCCAAGTTTGGCAGGCAAGTACACAAAGAAAGGAGGTATTAATCTTATCTCTCG	201926
QY	1606	TGATATATACGTGATGACAGAGAGATGATATGTAAAGCCCTCATATGGGCGCATGTGACTG	1665
Db	201927	TGATATATACCGTATATGAGGGGATGCTCTTATGTAATCTCCACATATGACCATAGCCATG	201986
QY	1666	GATTGAAAAAGATAGCCTTCTGATAAGAAAAAGTTGCACTCAAGCCTATACTAAAGA	1725
Db	201987	GATTAAAAAAGATAGTTTGTCTGAAGCTGAAAGAGCGGACGCCAGCCTATGCTAAAGA	202046
QY	1726	AAAAGGTATCTTACCTCCATCTCTCAGACGACAGATGTTAAAGCAAAATCCAACCTGGAATAG	1785
Db	202047	GAAAGGTTTGAACCCCTCTTCGACAGCAATTCAGGATTCGGAATACTGAGCAAAAAG	202106
QY	1786	TGCAGCAGCTTTTATACATGCTGTGAAGAGGGAAAAAGCAATTCACACTGTTGCACTGTC	1845
Db	202107	AGCAGAAAGCTATCTACAAACCGCGTGAAGAGAGCTTAAGAAAGCTCCCACTGATGCAATGCC	202166
QY	1846	ATATATGTTGAGCATCAAGCTTGAAGTAAAAAAAGTAAATTTATTAATTCCTCATAGA	1905
Db	202167	TTACAAATCTTAATTAATCTGTAAGAAAGTCAAAAAAGGTAAATTAATCAATCACTCATATATGA	202226
QY	1906	TCATTTACCATTAATTAATTTGCTTGTTTGTATGATCACACATTAACAAAGCTCCAAATGG	1965

Db	202227	CCATTACCATTAACATCAAAATTTGAACTGGTTTGACGAAAGCCCTTTTGTGAAGCACCTTAAGGG	202286
QY	1966	CTATACCTTTGGAAAGATTTGTTTGCAGCAGATTAACTACTACGTAGAACAACCCCTGACAGC	2025
Db	202287	GTATACCTCTTGAGATCTTTTGGCGACGTGTCAAGTACTATGTGCGAATCTCCAAACGAAACG	202346
QY	2026	TCCACATTTCTAATGATGATGGGCGCAATGCCAGTATGATGTGTTAAGGCAAGAAAGACCA	2085
Db	202347	TCCCGATTCAGATTAATGTTTGGTAAAGCTATGCGACCATGTGTTCAAAGAAACAAAATATGG	202406
QY	2086	CAGTCAATATCCAAATTAAGAATCTTCAAGCGGATGAGAG-----	2125
Db	202407	TCAAGCTGATACCAATCAACGGAAGAAACCAACGAGGAGAAACCTCAGACAGAAAAACC	202466
QY	2126	-----	2125
Db	202467	TGAGGAAGAAACCCCTCGAGAAAGAGAAACCGCAAGCGAAGAACCAAGATCTCCAAACC	202526
QY	2126	-----CCAGTAGAGAGAAACACCTGCTGTAGCCCAAGATCCCTCAAGTATGACATCTGA	2175
Db	202527	AACAGACGAACCCAGAAAGAAATCACACGAGGAAATCAGAAAGAACTCTCAGATCTGAACCTGA	202586
QY	2176	AAAAGTAGAAGCCCAACATCCAAAGAAAGAGAGAAATTTGCTTGGCAAGTAAGCATTTAG	2235
Db	202587	AAAGTTTAAAGAAAAACGAGAGAGAGCTGAGATTTACTTGGAAAAATTCAGATCCAT	202646
QY	2236	TCTGAAAGCCAAATGCACAGAAATCTTATGCTGTTCAGAAATTAATTTGACTCTTCAANT	2295
Db	202647	TATCAAGTCCAAATGCCAAAGAGACTCTCAGAGATTAATAAATAATTAATTTACTATTTTGGCAC	202706
QY	2296	TATGATTAACATATGATATCATGATGCGAGAAAGCGAAGAAATTAATCTGGCTGTAAAGGAAG	2355
Db	202707	CCAGACACAACATACTATTATTAATGCGAGAAAGCTGATAAATACTATTTGCTTTATTAAGAGAGAG	202766
QY	2356	TAA 2358	
Db	202767	TAA 202769	
RESULT 14			
ARI120270			
LOCUS	ARI120270	2290 bp	DNA linear PAT 16-MAY-2001
DEFINITION	Sequence 65 from patent US 6159469.		
ACCESSION	ARI120270		
VERSION	ARI120270.1	GI:14103846	
KEYWORDS			
SOURCE	Unknown.		
ORGANISM	Unknown.		
REFERENCE	1. (bases 1 to 2290)		
AUTHORS	Choi, S.H., Kunsch, C.A., Barrash, S.C., Dillon, P.J., Dougherty, B., Fannon, M.R. and Kosen, C.A.		
TITLE	Streptococcus pneumoniae antigens and vaccines		
JOURNAL	Patent: US 6159469-A 65 12-DEC-2000;		
FEATURES	Location/Qualifiers		
source	1..2290		
BASE COUNT	766 a 474 c 498 g 547 t	5 others	
ORIGIN			
Query Match	41.3%; Score 987.6; DB 6; Length 2290;		
Best Local Similarity	67.7%; Pred. No. 1e-200;		
Matches 1481:	Conservative 0; Mismatches 645; Indels 60; Gaps 5;		
QY	1	TTCTTACGATGGGACTGTATCAAGCTAGAGACGGTTAAGGAAA--TAAATCGTGTTC	57
Db	4	TTCCATATGACTTGGTGTCTCACCAAGCTGTCTCAAGGTAAAGAAAGTCTAATCGAGTTTC	63
QY	58	CTATATACATGTAACAAACAGACGACCAAAAAACGAGAAATTTGATCTCTGATGAGAGTTAG	117
Db	64	TTTATATACATGTTGTATGCGCTGTGTCAAAAGCAAGCAAACTATGACACCAAGATGAAGTCAG	123

OY	118	CAGGGGTAAAGGAAATTCATCTGTGACGAATAATCGTCAATCAABAATAACAGACCAGGCATATGT	177
Db	124	TAAAGGGAGGGGGATTCACAACGCCGCAAAATNGTNATCAAAATTACGATTAACGATTAAGTTATAT	183
OY	178	CACCTTCACATGGCGCCACTCATCTATTTATTCAAATGGTAAGTGCTCCTTATGACGCTATCAT	237
Db	184	GACCTCTATGGAGACCACTTATCTTTACTATTAATGGCAAGGTTCTTATATGATGCCATCAT	243
OY	238	CAGTGAAGAATTTACTCATGAAAGATCCAAAACCTAATAAGCTAAAAAGATGAGATATGTTAA	297
Db	244	CAGTGAAGAAGCTCCTCATGAAAGATCCGAATTTATACGTTGAAGATTCAGACATTGTCAA	303
OY	298	TGAGGTCAAGGGGTGGATATGTTATCAAGATGATGAGAAATACTATGTTTACTTTAAGA	357
Db	304	TGAATTCACAGGGGTGGTATGTTATCATTAAGGTAAACGGTAAATFACTATGTTATACCTTAAGA	363
OY	358	TGCTCCCAACCCGGATAAACGTCCCTCAAAAAGAGAAATCAATCGACAAAACAAAGAGA	417
Db	364	TGCAGCTCATGCGGATTAATTTGCGAACAAAAGAGATTTAAACGTCCAGACAGAGACG	423
OY	418	TATGTACACATCGTGAAGGTGGAACTCCAAGAAACGATGGTGCTGTTCCTTGGCACGTTC	477
Db	424	CAGTCAATATCAT-----AACTCAAGACGAGATTAATGCTGTTCGACGCCAGAC	474
OY	478	GCAAGAGCGCTATCTACAGATGATGTTATTTACTTTAATGCTTCGTGATATCATAGAGA	537
Db	475	CCMAAGACGTTATCAACGSGATGTGGGTATATCTTCAAAATGCAATCGATATCATTTAGAGA	534
OY	538	TACTGGTATGCTTATATATCGTTCCTCAAGAGATCATTACATTAATTCCTTAAGAATA	597
Db	535	CACGGGTATCTTATATCGTTCCTCAAGCGCACCATTTACATTAATTCCTTAAGAATA	594
OY	598	GTTATCACTAGCGAGTGTGGCTGCTGACAGAACCTTCTATCTGCTCGAGAAATCTGTTC	657
Db	595	GTTATCACTAGCGAGTGTGGCTGCTGACAGAACCT-----	629
OY	658	AAATTCAGAAGACTATGCGCGACAANAATACGATTAACACTTCAAGACAACTGGGTACC	717
Db	630	-----ATTGGAATGGGAAGCGGGGATCGTCTCTTCAAGTTCTTAAGTTAATAATAC	681
OY	718	TTCTGTAGCAATACAGAGACTACAAATTAATAACACAGAACAAACAGAACACTAAG	777
Db	662	AAATCAGCTCAACCAAGATTTGTCAAGAACCAACATGTGACTGTCACTCAACTTATCA	741
OY	778	TCAAGCAAGTCAAAATAATGACATGATGATGCTCTTGGAAACAGCTCTACAAACTGCCTT	837
Db	742	TCA---AAATCAAGGGGAAAAAATTTCAACCTTTACGTAATTTGATGTAACCTT	798
OY	838	GAGTCAACGACATGTAGAATCTATGAGCCCTTGTCTTGTATCCAGACAAATACAAATGCG	897
Db	799	ATCAAGAACGCATGTGAAATCTGATGGCTTATTTTTCGACCCAGCCAAATACAAAGTGC	858
OY	898	AACAGCTGAGAGGTGTGACAGTGCACACGAGAGATCAATACACTTATGCTTACTGTCA	957
Db	859	AACGCCACAGAGGTAGCTGTCTCTCATGTGTAACCATTTACCACTTATATCCCTTAAGAAC	918
OY	958	AATGTGGAATTTGGAAGAGAAATCGCTCTATATTTCCCTTCTGTTATCGTTCAACCA	1017
Db	919	AATGTGGAATTTGGAAGAGAAATGTGCTGTAATTTCCCTTCGTTATCGTTCAACCA	978
OY	1018	TTGGGTACAGATTCACAGGCCACAGAACCAACAGTCCACACACGACTCGGGAAGCTAGTCC	1077
Db	979	TTGGGTACCAATTTCAAGACCAGAACCAACCAATGCCACATCGACTCCGGAACCTAGTCC	1038
OY	1078	AGGCCCGCAACCTCCACCAAAATCTTAAATATGACTCAAAATCTTC-----TTT	1123
Db	1039	AAGTCCGCAACCTCCACCAAAATCTTCAACAGCTCCACAGCAATTCCAATTGATGAGAAAT	1098
OY	1126	GGTATGACAGCTGTGTACGAAAGTTGGGGAGAGATATATTTCGAAGAAAGGCGATC	1185
Db	1099	GGTCAAGAAAGCTTGTGCAAAAGTAGCGCATGTTATGTCTTTGAGAGAAATGAGAGTTTC	1158
OY	1186	TGCTTATGTCTTTCGAAAGATTTTACATCTGAACCTGTTTAAAAATCTTGAAGCAAGT	1245

[illegible]

RESULT 15	AX343074	2639 bp	DNA	linear	PAT 12-JAN-2002
LOCUS	AX343074	Sequence 5 from Patent WO019834.			
DEFINITION					
ACCESSION	AX343074				





GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 16, 2002, 23:10:35 ; Search time 340 Seconds

(without alignments)  
15823.600 Million cell updates/sec

Title: US-08-961-083-55

Perfect score: 2389

Sequence: 1 TTCTTACGAGTTGGACGTCT.....TAAGTAAGCAAAAATAAAC 2389

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Lasting first 45 summaries

Database : N\_Geneseq\_101002:\*

1: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:\*

2: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:\*

3: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:\*

4: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:\*

5: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:\*

6: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:\*

7: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:\*

8: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:\*

9: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:\*

10: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:\*

11: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:\*

12: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:\*

13: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:\*

14: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:\*

15: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:\*

16: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:\*

17: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:\*

18: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:\*

19: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:\*

20: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:\*

21: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:\*

22: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:\*

23: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:\*

24: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2388	100.0	2389	19	AAV27351	Streptococcus pneu
2	2388	100.0	2389	24	AB084819	S. pneumoniae SP03
3	2388	100.0	2451	21	AAA47604	Recombinant varian
4	2388	100.0	8195	19	AAV52227	Streptococcus pneu
5	1374.2	57.5	2523	21	AAA65731	Streptococcus pneu
6	1374.2	57.5	2647	21	AAA65736	Streptococcus pneu
7	1374.2	57.5	2647	24	ABK15103	DNA encoding Strep
8	1011.8	42.0	2478	21	AAA08557	S. pneumoniae 92 k
9	1003.8	42.0	2481	21	AAA05417	Streptococcus pneu

10	991	41.5	2531	21	AAA47605	Recombinant varian
11	990	41.4	2531	21	AAA47602	Recombinant varian
12	987.6	41.3	2290	19	AAV27356	Streptococcus pneu
13	987.6	41.3	2290	24	AB084824	S. pneumoniae SP04
14	980.8	41.1	2639	21	AAA65737	Streptococcus pneu
15	980.8	41.1	2639	24	ABK15104	DNA encoding Strep
16	869.4	36.4	2163	20	AAV52394	Streptococcus pneu
17	653.8	27.4	2359	19	AAV52376	Streptococcus pneu
18	548.4	23.0	3171	21	AAA65739	Streptococcus pneu
19	385.4	16.1	1342	19	AAV27414	Streptococcus pneu
20	385.4	16.1	1342	24	AB084882	S. pneumoniae SP10
21	385.4	16.1	1455	21	AAA65733	Streptococcus pneu
22	385.4	16.1	1455	21	AAA47603	Recombinant varian
23	385.4	16.1	1455	21	AAA05473	Streptococcus pneu
24	385.4	16.1	1455	21	AAZ91804	Streptococcus pneu
25	385.4	16.1	3120	21	AAA65730	Streptococcus pneu
26	385.4	16.1	5048	21	AAA65735	Streptococcus pneu
27	385.4	16.1	5048	24	ABK15101	DNA encoding Strep
28	385.4	16.1	6867	19	AAV52325	Streptococcus pneu
29	381.4	16.0	973	19	AAV52488	Streptococcus pneu
30	347	14.5	2528	21	AAA65738	Streptococcus pneu
31	347	14.5	2528	24	ABK15105	DNA encoding Strep
32	247.4	10.4	504	21	AAA08556	S. pneumoniae 20 k
33	243.4	10.2	1684	19	AAV52391	Streptococcus pneu
34	214.6	9.0	2475	24	ABN66838	Streptococcus poly
35	214.6	9.0	2478	22	AA500036	Streptococcus pyog
36	211.4	8.8	1146	21	AAA05814	Group B Streptococ
37	211.4	8.8	2466	24	ABN69535	Streptococcus poly
38	211.4	8.8	2466	24	ABN70334	Streptococcus poly
39	211.4	8.8	2469	21	AAA05811	Group B Streptococ
40	211.4	8.8	2469	22	AA500038	Streptococcus agal
41	209.8	8.8	2469	21	AAA65740	Streptococcus pneu
42	209.8	8.8	5215	20	AAK91105	Group B Streptococ
43	206.6	8.6	2472	21	AAA65741	Streptococcus pneu
44	201.6	8.4	1398	24	ABN66839	Streptococcus poly
45	201.2	8.4	819	24	ABK15106	DNA encoding Strep

## ALIGNMENTS

RESULT 1  
AAV27351  
ID AAV27351 standard; DNA; 2389 BP.  
AC AAV27351:  
DT 02-OCT-1998 (first entry)  
DE Streptococcus pneumoniae SP0036 nucleotide.  
XX  
KW Streptococcus pneumoniae: antigen; vaccine; infection; diagnosis;  
KW detection; pneumonia; otitis media; meningitis; ss.  
XX  
OS Streptococcus pneumoniae.  
XX  
FH Key Location/Qualifiers  
FT CDS 2..2389  
FT /tag= a  
FT /product= "SP0036"  
FT /transl\_except= (pos:1367..1369,aa:Xaa)  
FT /note= "no stop codon given; Xaa is unspecified"  
PN W09818930-A2.  
XX  
PD 07-MAY-1998.  
XX  
PF 30-OCT-1997; 97WO-US19422.  
XX  
PR 31-OCT-1996; 96US-0029960.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.

PI Choi GH, Hromockyj A, Johnson LS, Kunsch CA;  
XX MPI: 1998-272224/24.  
DR P-PSDB; AAM55090.  
XX  
XX Nucleic acid encoding antigenic peptide(s) from Streptococcus  
PT pneumoniae - or their epitope-containing fragments, useful in  
PT protective or therapeutic vaccines, and for diagnosis  
PS  
XX Claim 1: Page 59; 118bp; English.  
XX  
XX The present sequence encodes a protein from Streptococcus pneumoniae.  
CC The nucleic acid sequence encoding the Streptococcus pneumoniae protein  
CC can be useful in vaccines for inducing protective antibodies against  
CC Streptococcus pneumoniae, for treatment or prevention of infection e.g.  
CC pneumonia, otitis media or meningitis. Probes based on the nucleic acid  
CC are used to detect Streptococcus infection (by usual hybridisation or  
CC amplification methods), also for isolating Streptococcus genes or their  
CC allelic variants. The protein can be used similarly for detecting specific  
CC antibodies in standard immunoassays, especially for diagnosing or  
CC monitoring infections. Antibodies which bind the protein are used to  
CC immunisation (optionally coupled to a toxin). Vaccines are administered,  
CC e.g. by injection, orally or through the skin, typically at 0.01-1000  
CC (especially 10-300) mu g/ml per dose.  
XX  
XX Sequence 2389 BP; 830 A; 461 C; 486 G; 611 T; 1 other:  
SQ  
Query Match 100.0%; Score 2388; DB 19; Length 2389;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2389; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

|||||  
Db ATCAGCTAGAGAGTTGGCTGTCAGAGAGCCTTCTATCTGGTGGAGGAATCTGTCAAA 660  
QY 601 TTCAAAACCTATCGCCGACAAAATATGCGATTAACATCTTAAGAACAACTGGGTACTTC 720  
QY 661 TTCAAAACCTATCGCCGACAAAATATGCGATTAACATCTTAAGAACAACTGGGTACTTC 720  
Db 661 TTCAAAACCTATCGCCGACAAAATATGCGATTAACATCTTAAGAACAACTGGGTACTTC 720  
QY 721 TGTAAACATTCAGAACTACAAATATCTAAACACCAACAGCAGCACTAACAGTCA 780  
Db 721 TGTAAACATTCAGAACTACAAATATCTAAACACCAACAGCAGCACTAACAGTCA 780  
QY 781 AGCAAGTCAAAGTATGACATTTGATAGTCTTTGAAAAGCTCTTACAAAGCTTTGAG 840  
Db 781 AGCAAGTCAAAGTATGACATTTGATAGTCTTTGAAAAGCTCTTACAAAGCTTTGAG 840  
QY 841 TCACACACATGTATGATCTGATGCGCTTGCTCTTTGATCCAGACAAATACAAAGTGCAG 900  
Db 841 TCACACACATGTATGATCTGATGCGCTTGCTCTTTGATCCAGACAAATACAAAGTGCAG 900  
QY 901 AGCTAGAGTGTGTCAGTGCACACAGAGATCATTCACACTTCATCCCTTACTCTCAAT 960  
Db 901 AGCTAGAGTGTGTCAGTGCACACAGAGATCATTCACACTTCATCCCTTACTCTCAAT 960  
QY 961 GTCTGAATTTGGAAGAGCAATGCGCTGTATTTATCCCTTCCTGTTATCGTTCAACCATTG 1020  
Db 961 GTCTGAATTTGGAAGAGCAATGCGCTGTATTTATCCCTTCCTGTTATCGTTCAACCATTG 1020  
QY 1021 GGTACCAAGTTCAAGGCCAGAACACCAACAGTCCACACAGCATCTCGGAACCTATGTCAG 1080  
Db 1021 GGTACCAAGTTCAAGGCCAGAACACCAACAGTCCACACAGCATCTCGGAACCTATGTCAG 1080  
QY 1081 CCGGCAACCTGCACCAATCTTAATATAGCTCAATCTCTGTTGTTAGTACAGTGT 1140  
Db 1081 CCGGCAACCTGCACCAATCTTAATATAGCTCAATCTCTGTTGTTAGTACAGTGT 1140  
QY 1141 ACGAAAAGTTGGGGAAGATATGTATTCGAAGAAAAGGCACTCTGTTATGCTTTGC 1200  
Db 1141 ACGAAAAGTTGGGGAAGATATGTATTCGAAGAAAAGGCACTCTGTTATGCTTTGC 1200  
QY 1201 GAAAGATTTACCACTGGAAGCTTTAAATCTGGAACCAAGTTATCAAAAACGAAGAG 1260  
Db 1201 GAAAGATTTACCACTGGAAGCTTTAAATCTGGAACCAAGTTATCAAAAACGAAGAG 1260  
QY 1261 TGTTCACACACTTAACTGCTAAAGAAAGAAATGTTGCTCGTGCAGCAAGATTTTA 1320  
Db 1261 TGTTCACACACTTAACTGCTAAAGAAAGAAATGTTGCTCGTGCAGCAAGATTTTA 1320  
QY 1321 TGTATTAAGCATATATCTGTTAACTGAGGCTCATAAAGCCTTGTGNAATPAAGGTCG 1380  
Db 1321 TGTATTAAGCATATATCTGTTAACTGAGGCTCATAAAGCCTTGTGNAATPAAGGTCG 1380  
QY 1381 TAATTCGTATTTCCAAAGCTTATAGCAAAATTTATAGAACCTTGAATGATCCACTTA 1440  
Db 1381 TAATTCGTATTTCCAAAGCTTATAGCAAAATTTATAGAACCTTGAATGATCCACTTA 1440  
QY 1441 TAAAGAAAATTTGATGATGATTTATGTCATTCTGAGACCAATTAACCATCCAGAGCG 1500  
Db 1441 TAAAGAAAATTTGATGATGATTTATGTCATTCTGAGACCAATTAACCATCCAGAGCG 1500  
QY 1501 ACTTGGCAAAACCAATTTCTCAAAATGAGTATCTGAAAGCAAGTTCGATTTGCTCAAT 1560  
Db 1501 ACTTGGCAAAACCAATTTCTCAAAATGAGTATCTGAAAGCAAGTTCGATTTGCTCAAT 1560  
QY 1561 AGCTGATTAAGTATACACGTCAGATGCTTACATTTTGTATGAACATGATATATACGTA 1620  
Db 1561 AGCTGATTAAGTATACACGTCAGATGCTTACATTTTGTATGAACATGATATATACGTA 1620  
QY 1621 TGAAGAGATGATATGTATACGCTCATATGAGCCATATGCTGATGTTGAAAGATAG 1680  
Db 1621 TGAAGAGATGATATGTATACGCTCATATGAGCCATATGCTGATGTTGAAAGATAG 1680  
QY 1681 CCTTCTGATTAAGAAAAGTTGCAGCTCAAGCTTATATCTAATAAGAAAAGGATCTCTACC 1740  
|||||

Db 1681 CTTTTCTGATTAAGAAAAAGTTGACCTCAAGCTATTAAGAAAAAGTATCCACC 1740  
QY 1741 TCCATCTCCAGACGAGATGTAAAGCAATCCAACTGAGATAGTGCAGCAGCTATTTA 1800  
Db 1741 TCCATCTCCAGACGAGATGTAAAGCAATCCAACTGAGATAGTGCAGCAGCTATTTA 1800  
QY 1801 CAATCGTGTAAAGGGGAAAAAGAAATTCACCTGCTGCACTTCATATATGTTGAGCA 1860  
Db 1801 CAATCGTGTAAAGGGGAAAAAGAAATTCACCTGCTGCACTTCATATATGTTGAGCA 1860  
QY 1861 TACAGTTGAGGTTAAAAAGCGTAATTTGATTTATCTCATTAAGATCATTAATAT 1920  
Db 1861 TACAGTTGAGGTTAAAAAGCGTAATTTGATTTATCTCATTAAGATCATTAATAT 1920  
QY 1921 TAAATTTGCTGGTTTGATGATACACATACAAAGCTCCAAATGCTATACCTTTGAGA 1980  
Db 1921 TAAATTTGCTGGTTTGATGATACACATACAAAGCTCCAAATGCTATACCTTTGAGA 1980  
QY 1981 TTTGTTTGCAGCATTTAAGTACTAGTAAACACCCCTGACGAACGTCACATTCCTAATGA 2040  
Db 1981 TTTGTTTGCAGCATTTAAGTACTAGTAAACACCCCTGACGAACGTCACATTCCTAATGA 2040  
QY 2041 TGGATGGGGGAATGCGCATGATGATGTTAGGCAAGAAAGACACAGTAAATCCAAA 2100  
Db 2041 TGGATGGGGGAATGCGCATGATGATGTTAGGCAAGAAAGACACAGTAAATCCAAA 2100  
QY 2101 TAAAGACTTCAAAAGCGGATGAAGAGCCAGTAAAGAAACACCTGCTGAGCAGAACTCCC 2160  
Db 2101 TAAAGACTTCAAAAGCGGATGAAGAGCCAGTAAAGAAACACCTGCTGAGCAGAACTCCC 2160  
QY 2161 TCAAGTAGAGACTGAAAAAGTAGAAGCCCAACTCAAAAGAAAGCAAGAAATTTGCTGGCAA 2220  
Db 2161 TCAAGTAGAGACTGAAAAAGTAGAAGCCCAACTCAAAAGAAAGCAAGAAATTTGCTGGCAA 2220  
QY 2221 AGTAAGGAGATTTAGTCTGGAAGCCATGCAACAGAAACTGACTGCTTTAGCAATTA 2280  
Db 2221 AGTAAGGAGATTTAGTCTGGAAGCCATGCAACAGAAACTGACTGCTTTAGCAATTA 2280  
QY 2281 TTTGACTCTTCAATTAATGATTAACAATAGTATCATATGCGCAAGAAACATACTTGC 2340  
Db 2281 TTTGACTCTTCAATTAATGATTAACAATAGTATCATATGCGCAAGAAACATACTTGC 2340  
QY 2341 GTTGTAAAAAGAAATATCTTCAATCTGTAAGTAAAGAAAAATAAAC 2389  
Db 2341 GTTGTAAAAAGAAATATCTTCAATCTGTAAGTAAAGAAAAATAAAC 2389  
RESULT 2  
ABQ84819  
ID ABQ84819 standard: DNA; 2389 BP.  
XX  
AC ABQ84819;  
XX  
DT 04-SEP-2002 (first entry)  
XX  
DE S. pneumoniae SP036 nucleotide sequence SEQ ID NO:55.  
XX  
KM Streptococcus pneumoniae: epitope; vaccine; antigenic protein;  
XX antibacterial; Streptococcal infection; detection; gene; ds.  
XX  
OS Streptococcus pneumoniae.  
XX  
PN US2002061545-A1.  
PD  
XX 23-MAY-2002.  
XX  
PF 22-JAN-2001; 2001US-0765272.  
XX  
PR 30-OCT-1997; 97US-0961083.  
XX  
PA (CHOI/) CHOI G H.  
PA (KUNSH/) KUNSH C A.  
PA (BARA/) BARASH S C.

PA (DILL/) DILLON P J.  
PA (DOUG/) DOUGHERTY B.  
PA (FANN/) FANNON M R.  
PA (ROSE/) ROSEN C A.  
PI Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;  
PI Rosen CA;  
DR WPI: 2002-479261/51.  
DR P-PSDB: ABP54584.  
XX  
XX  
XX New Streptococcus pneumoniae antigens, useful for detecting  
XX Streptococcus and for preventing or attenuating disease caused by  
XX Streptococcus infection -  
PS Claim 1; Page 27; 70pp; English.  
XX  
XX  
XX ABQ84792 to ABQ84904 represents nucleic acids which encode the  
XX Streptococcus pneumoniae antigens given in ABP54557 to ABP54669.  
XX The S. pneumoniae antigens have antibacterial activity and can be  
XX used in vaccines. The S. pneumoniae antigens can also be used to  
XX prevent or attenuate a Streptococcal infection in an animal. The  
XX polynucleotides encoding the S. pneumoniae antigens can be used to  
XX detect Streptococcus nucleic acids. ABQ84905 to ABQ85130 represent  
XX primers used in the cloning of S. pneumoniae ORFs (open reading frames)  
XX which are used in an example from the present invention.  
SQ Sequence 2389 BP; 830 A; 461 C; 486 G; 611 T; 1 other;  
Query Match 100.0%; Score 2388; DB 24; Length 2389;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2389; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TTTCTTACGATTTGGGAGTGTATCAAGCTAGACGCGTTAAGAAAAATATGTTCTTA 60  
Db 1 TTTCTTACGATTTGGGAGTGTATCAAGCTAGACGCGTTAAGAAAAATATGTTCTTA 60  
QY 61 TATGATGAGAAAAACAGCGCAAAAAAGGAGAAATTTGACCTGATGAGGTTAGCAA 120  
Db 61 TATGATGAGAAAAACAGCGCAAAAAAGGAGAAATTTGACCTGATGAGGTTAGCAA 120  
QY 121 GCGTGAAGGAATCAATGCTGAGCAAAATGCTATCAAGATTAACAGACCAAGCTATGTCAC 180  
Db 121 GCGTGAAGGAATCAATGCTGAGCAAAATGCTATCAAGATTAACAGACCAAGCTATGTCAC 180  
QY 181 TTTCAATGCGGACCATATCATTTTACAAATGTAAGGTTCTTATAGCGTATCATCAG 240  
Db 181 TTTCAATGCGGACCATATCATTTTACAAATGTAAGGTTCTTATAGCGTATCATCAG 240  
QY 241 TGAAGATTTACTCATGAAAGATCCAACTTAAGCTAAAGATGAGATTTGTTATGA 300  
Db 241 TGAAGATTTACTCATGAAAGATCCAACTTAAGCTAAAGATGAGATTTGTTATGA 300  
QY 301 GGTCAAGGATGATATGTTTCAAGGTAGATGAAAAATATATCTTATACCTTAAGATGC 360  
Db 301 GGTCAAGGATGATATGTTTCAAGGTAGATGAAAAATATATCTTATACCTTAAGATGC 360  
QY 361 TGCCCAACCGCGATACCTCGTACAAAAGAGAAATCAATCGACAAAACAAGGCAATAG 420  
Db 361 TGCCCAACCGCGATACCTCGTACAAAAGAGAAATCAATCGACAAAACAAGGCAATAG 420  
QY 421 TCAACATCGTGAAGGTGGAATCCCAAGAAAGATGAGTGGTGGCTTGACAGTTGCA 480  
Db 421 TCAACATCGTGAAGGTGGAATCCCAAGAAAGATGAGTGGTGGCTTGACAGTTGCA 480  
QY 481 AGGACGCTATACACAGATGATGTTATCTTAAATGCTTGTGATTCATAGAGATAC 540  
Db 481 AGGACGCTATACACAGATGATGTTATCTTAAATGCTTGTGATTCATAGAGATAC 540  
QY 541 TGGTGAATGCTTATATCTTCTCATGAGATCATTTACCATTAATTCCTTAAGATGATT 600  
Db 541 TGGTGAATGCTTATATGCTTCTCATGAGATCATTTACCATTAATTCCTTAAGATGATT 600



QY 601 ATCAGCTAGCGAGTGGCTGCTGTCAGAGAGCCTTCATCTGTCGTGAGAAATCTGTCAA 660  
Db 601 ATCAGCTAGCGAGTGGCTGCTGTCAGAGAGCCTTCATCTGTCGTGAGAAATCTGTCAA 660  
QY 661 TTCAAGACATCTGCGCGACAAATAGCGATACACTTCAAGAACAACTGGGTACTTC 720  
Db 661 TTCAAGACATCTGCGCGACAAATAGCGATACACTTCAAGAACAACTGGGTACTTC 720  
QY 721 TGTAAAGCAATCCAGAGACTCAATCTAACAAGACAGACAGCAACTAAGTCA 780  
Db 721 TGTAAAGCAATCCAGAGACTCAATCTAACAAGACAGACAGCAACTAAGTCA 780  
QY 781 AGCAAGTCAAAAGTAATGACATTTAGTCTCTTGAACAGCTCTAGCAACTGGCTTTGAG 840  
Db 781 AGCAAGTCAAAAGTAATGACATTTAGTCTCTTGAACAGCTCTAGCAACTGGCTTTGAG 840  
QY 841 TCACAGCAGTGTGAGTGTGATGCGCTGTCTTGTGATCCAGCAAAATCAAGTGAAC 900  
Db 841 TCACAGCAGTGTGAGTGTGATGCGCTGTCTTGTGATCCAGCAAAATCAAGTGAAC 900  
QY 901 AGCTAGAGGTGTCAGTGGCCACAGGAGATCTACCACTTCCTTACTCTCAAT 960  
Db 901 AGCTAGAGGTGTCAGTGGCCACAGGAGATCTACCACTTCCTTACTCTCAAT 960  
QY 961 GTCTGATTTGGAAGAAAGAAATGCTGTATTTCCCTTGTGATCTTCAACCAATG 1020  
Db 961 GTCTGATTTGGAAGAAAGAAATGCTGTATTTCCCTTGTGATCTTCAACCAATG 1020  
QY 1021 GGTACCAAGTTCAGAGCCAGAACCAACAGTCCACAAACGACCCGGAACCTAGTCCAG 1080  
Db 1021 GGTACCAAGTTCAGAGCCAGAACCAACAGTCCACAAACGACCCGGAACCTAGTCCAG 1080  
QY 1081 CCCGCAACCTGACCAAAATCTTAAATAGACTCAATTTCTTGTGATGTCAGTGTG 1140  
Db 1081 CCCGCAACCTGACCAAAATCTTAAATAGACTCAATTTCTTGTGATGTCAGTGTG 1140  
QY 1141 ACGAAAAGTTGGGGAAGATATGATTCGAAGAAAAGGCGATCTCTGTTATGCTTTG 1200  
Db 1141 ACGAAAAGTTGGGGAAGATATGATTCGAAGAAAAGGCGATCTCTGTTATGCTTTG 1200  
QY 1201 GAAGAATTTACCATCTGAAACCTGTAAATCTTGAAGAGCAATTCAAACAAAGAG 1260  
Db 1201 GAAGAATTTACCATCTGAAACCTGTAAATCTTGAAGAGCAATTCAAACAAAGAG 1260  
QY 1261 TGTTCACACACTTAACTGCTTAAAAAGAAATGTTGCTCTGTCGACCAAGATTTTA 1320  
Db 1261 TGTTCACACACTTAACTGCTTAAAAAGAAATGTTGCTCTGTCGACCAAGATTTTA 1320  
QY 1321 TGTAAAGCATTAATCTGTTAACTGAGGCTCAATAAGCCTTGTGNAATTAAGGCTG 1380  
Db 1321 TGTAAAGCATTAATCTGTTAACTGAGGCTCAATAAGCCTTGTGNAATTAAGGCTG 1380  
QY 1381 TAATTCGATTTCCAGACCTTACAGCAATTAATAGAACGCTTGAAGATGAAATGACTAA 1440  
Db 1381 TAATTCGATTTCCAGACCTTACAGCAATTAATAGAACGCTTGAAGATGAAATGACTAA 1440  
QY 1441 TAAAGAAAAATTTGATAGATGATTTATTTGCAATCTAGACCAATTAACCAAGCG 1500  
Db 1441 TAAAGAAAAATTTGATAGATGATTTATTTGCAATCTAGACCAATTAACCAAGCG 1500  
QY 1501 ACTTGGCAACCAATTTCTCAATTTGAGTACTGAAGAGCAAGTTCGATTTGCTCAAT 1560  
Db 1501 ACTTGGCAACCAATTTCTCAATTTGAGTACTGAAGAGCAAGTTCGATTTGCTCAAT 1560  
QY 1561 AGCTGATTAAGTATACAGCTCAGATGTTTCATTTTGTGATGAACATGATTAATCACTGA 1620  
Db 1561 AGCTGATTAAGTATACAGCTCAGATGTTTCATTTTGTGATGAACATGATTAATCACTGA 1620  
QY 1621 TGAAGAGATGATATGATTAAGCGCTATATGCGCATAGTCACTGATGATGGAAGATAG 1680  
Db 1621 TGAAGAGATGATATGATTAAGCGCTATATGCGCATAGTCACTGATGATGGAAGATAG 1680  
QY 1681 CCTTCTGATTAAGAAAAAGTTGACCTCAAGCCTTACTTAAAGAAAAAGTATCTTACC 1740

Db 1681 CCTTCTGATTAAGAAAAAGTTGACCTCAAGCCTTACTTAAAGAAAAAGTATCTTACC 1740  
QY 1741 TCCATCCAGACGAGATGTTAAACCAATCCAACTGAGATAGTGCAGCGACTTTTA 1800  
Db 1741 TCCATCCAGACGAGATGTTAAACCAATCCAACTGAGATAGTGCAGCGACTTTTA 1800  
QY 1801 CAATCGTGTGAAGGGGAAAAAACGAATTCACATGCTGTCGATTCATTAATGTTGAGCA 1860  
Db 1801 CAATCGTGTGAAGGGGAAAAAACGAATTCACATGCTGTCGATTCATTAATGTTGAGCA 1860  
QY 1861 TACAGTTGAGGTTAAAAACGATTAATTTGATTTCTCATTAAGATCATTTACATTAAT 1920  
Db 1861 TACAGTTGAGGTTAAAAACGATTAATTTGATTTCTCATTAAGATCATTTACATTAAT 1920  
QY 1921 TAAATTTGCTGTTGATGATACACATTAACAAGTCCAAATGGCTATACCTTGAAGA 1980  
Db 1921 TAAATTTGCTGTTGATGATACACATTAACAAGTCCAAATGGCTATACCTTGAAGA 1980  
QY 1981 TTTGTTGGGACGATTAAGTACTAGCTAGACACCTGACGAACTCCATTTCTAATGA 2040  
Db 1981 TTTGTTGGGACGATTAAGTACTAGCTAGACACCTGACGAACTCCATTTCTAATGA 2040  
QY 2041 TGGATGGGCAATGCGACATGAGCATGTTAGGCAAGAAAGACACAGTGAAGATCAAA 2100  
Db 2041 TGGATGGGCAATGCGACATGAGCATGTTAGGCAAGAAAGACACAGTGAAGATCAAA 2100  
QY 2101 TAAAGCTTCAAAAGGGATGAAGAGCCATGAGGAAACACCTGTCGACCGAAATGCC 2160  
Db 2101 TAAAGCTTCAAAAGGGATGAAGAGCCATGAGGAAACACCTGTCGACCGAAATGCC 2160  
QY 2161 TCAAGTAGAGACTGAAAGAAAGTAGAACCCCACTCAAGAAAGCAAGATTTGCTTCGAA 2220  
Db 2161 TCAAGTAGAGACTGAAAGAAAGTAGAACCCCACTCAAGAAAGCAAGATTTGCTTCGAA 2220  
QY 2221 AGTAAGGATTTAGTCTGAAAGCCAAATGCAACAGAAACTTACTGCTTTACGAATTA 2280  
Db 2221 AGTAAGGATTTAGTCTGAAAGCCAAATGCAACAGAAACTTACTGCTTTACGAATTA 2280  
QY 2281 TTTGACTCTTCAATTAATGATTAACAATGATATCATGCGAGAAAGCAAGAAATTAATTC 2340  
Db 2281 TTTGACTCTTCAATTAATGATTAACAATGATATCATGCGAGAAAGCAAGAAATTAATTC 2340  
QY 2341 GTTGTAAAAAGGAAGTAATCTTCATCTGTAAGTAAGAAAAAATAAAC 2389  
Db 2341 GTTGTAAAAAGGAAGTAATCTTCATCTGTAAGTAAGAAAAAATAAAC 2389  
  
RESULT 3  
AAA47604  
ID AAA47604 standard; DNA; 2451 BP.  
XX  
AC AAA47604;  
XX  
DT 20-OCT-2000 (first entry)  
XX  
DE Recombinant variant of Sp36 gene (Sp36A) of S. pneumoniae.  
XX  
KW Streptococcus pneumoniae; infection; vaccine; coiled coil region;  
XX  
KW histidine triad residue; Sp36; antibody; otitis media;  
KW nasopharyngeal infection; bronchial infection; bronchitis; sepsis;  
XX  
KW meningitis; lobar pneumonia; ds.  
XX  
OS Streptococcus pneumoniae.  
XX  
FH key Location/Qualifiers  
FT 1..2451  
FT CDS /\*tag= a  
FT /product= Sp36a polypeptide  
PN MO200037105-A2.  
XX  
PD 29-JUN-2000.

XX 21-DEC-1999; 99WO-US30390.  
PF  
XX 21-DEC-1998; 98US-0113048.  
PR  
XX  
XX (MED1-) MEDIMUNE INC.  
PA  
PI Johnson LS, Koenig S, Adamou JE;  
XX WPI; 2000-452129/39.  
XX P-SDB; AAB01468.  
DR  
XX  
PT Vaccine useful for prophylaxis and treatment of pneumococcal infections  
PT such as otitis media, nasopharyngeal and bronchial infections,  
PT comprises Streptococcus pneumoniae proteins  
XX  
XX  
PS Disclosure; Page 64-65; 70pp; English.  
XX  
XX Although a number of proteins have been suggested as being involved  
CC in the pathogenicity of Streptococcus pneumoniae, there still remains  
CC a need to identify polypeptides having epitopes in common from  
CC various strains of S. pneumoniae in order to utilize such  
CC polypeptides in vaccines to protect against a wide variety of  
CC S. pneumoniae. New vaccine compositions are described which comprise a  
CC Streptococcus pneumoniae polypeptide (or fragments) of 80 - 680 amino  
CC acids in length that comprise at least one histidine triad residue  
CC (HxxHxH) or a coiled-coil region, or an antibody directed against  
CC these features. The vaccine is useful in protecting against infection  
CC by Streptococcus pneumoniae. The vaccine composition comprising  
CC antibodies to is useful for passive immunization for treating  
CC pneumococcal infections which includes otitis media, nasopharyngeal  
CC and bronchial infections.  
XX  
XX Sequence 2451 BP; 849 A; 467 C; 499 G; 635 T; 1 other;  
SO  
Query Match 100.0%; Score 2388; DB 21; Length 2451;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2389; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 TTCTTACAGTGGGACGTGTATCTACAGCTAGACGGTTAAGAAATTAATGTTTCTTA 60  
DB 60 TTCTTACAGTGGGACGTGTATCTACAGCTAGACGGTTAAGAAATTAATGTTTCTTA 119  
OY 61 TATAGATGAAAAACAAACGACGCAAAAAACGAGAAATTGACTCTGATGAGTTAGCAA 120  
DB 120 TATAGATGAAAAACAAACGACGCAAAAAACGAGAAATTGACTCTGATGAGTTAGCAA 179  
OY 121 GCGTGAAGGAATCAATGCTGAGCAAAATCGTCATCAAGATAACAGACCAGCTATGTCCAC 180  
DB 180 GCGTGAAGGAATCAATGCTGAGCAAAATCGTCATCAAGATAACAGACCAGCTATGTCCAC 239  
OY 181 TTTCACATGGGACACATATCATATTATTCAAATGTAAGTTCCTTATAGCGTATACACAG 240  
DB 240 TTTCACATGGGACACATATCATATTATTCAAATGTAAGTTCCTTATAGCGTATACACAG 299  
OY 241 TGAAGAAATTAATCTATGAAAGATCCAAACTATAAGCTAAAGATGAGGATATTGTTAATGA 300  
DB 300 TGAAGAAATTAATCTATGAAAGATCCAAACTATAAGCTAAAGATGAGGATATTGTTAATGA 359  
OY 301 GGTCAAGGGTGGATGTTTATCAAGGTAGATGCAAAATTAATGTTTAACTTAAGATGC 360  
DB 360 GGTCAAGGGTGGATGTTTATCAAGGTAGATGCAAAATTAATGTTTAACTTAAGATGC 419  
OY 361 TGGCCACGGGGATTAAGTTCGTACAAAGAGAAATCAATCGCAAAAAACAGAGCATAG 420  
DB 420 TGGCCACGGGGATTAAGTTCGTACAAAGAGAAATCAATCGCAAAAAACAGAGCATAG 479  
OY 421 TCAACATCGTGAAGGTGAATCCCAAGAAAGATGCTGCTTGGCCTTGACAGCTTGCA 480  
DB 480 TCAACATCGTGAAGGTGAATCCCAAGAAAGATGCTGCTTGGCCTTGACAGCTTGCA 539  
OY 481 AGGACGTATACATACAGATATGTTTATATCTTTAATGCTTGTGATATCATAGAGATAC 540  
DB 481 AGGACGTATACATACAGATATGTTTATATCTTTAATGCTTGTGATATCATAGAGATAC 540

DB 540 AGGACGTATACATACAGATATGTTTATATCTTTAATGCTTGTGATATCATAGAGATAC 599  
OY 541 TGGTATGCTTATATCGTTCCTCATGAGATCATTTACCATTAATTCCTAAGAAATGAGTT 600  
DB 600 TGGTATGCTTATATCGTTCCTCATGAGATCATTTACCATTAATTCCTAAGAAATGAGTT 659  
OY 601 ATCAGTACGAGTTGGCTGCTGAGAAAGCCCTTCATCTGTGTGAGAAATCTGTCAA 660  
DB 660 ATCAGTACGAGTTGGCTGCTGAGAAAGCCCTTCATCTGTGTGAGAAATCTGTCAA 719  
OY 661 TTCAAGAACCTATATGCGCGACAAAAATAGCATTAACATTTCAAGAAACAATCGTACCTTC 720  
DB 720 TTCAAGAACCTATATGCGCGACAAAAATAGCATTAACATTTCAAGAAACAATCGTACCTTC 779  
OY 721 TGTAAAGCATTCAGACACTCAAAATTAACACACAGCAACAGCAACACTAACAGTCA 780  
DB 780 TGTAAAGCATTCAGACACTCAAAATTAACACACAGCAACAGCAACACTAACAGTCA 839  
OY 781 AGCAAGTCAAAAGTAATGACATTTGATAGTCTCTTGAACACGCTTACAAAGTCCCTTGAG 840  
DB 840 AGCAAGTCAAAAGTAATGACATTTGATAGTCTCTTGAACACGCTTACAAAGTCCCTTGAG 899  
OY 841 TCAACGACATGTAGATCTGATGCGCTTGTCTTGTATCCAGACACAAATCACAGTGCAG 900  
DB 900 TCAACGACATGTAGATCTGATGCGCTTGTCTTGTATCCAGACACAAATCACAGTGCAG 959  
OY 901 AGCTAGAGGTTGTCAGTGCACACGAGAGATCATACACTTCCTACCTTACCTCAAT 960  
DB 960 AGCTAGAGGTTGTCAGTGCACACGAGAGATCATACACTTCCTACCTTACCTCAAT 1019  
OY 961 GTCGATTTGGAAGAACGAATTCGCTGTAATTAATTCCTTGTATGCTTCAACCAATG 1020  
DB 1020 GTCGATTTGGAAGAACGAATTCGCTGTAATTAATTCCTTGTATGCTTCAACCAATG 1079  
OY 1021 GGTACCGATTTCAAGCCGACAGAACCAAGTCCAAACGACTCCGAACTTACCCAGG 1080  
DB 1080 GGTACCGATTTCAAGCCGACAGAACCAAGTCCAAACGACTCCGAACTTACCCAGG 1139  
OY 1081 CCGGCAACCTGACCAATATCTTAATAATAGACTCAATCTTCTTGTGTGTGCTGCTGCT 1140  
DB 1140 CCGGCAACCTGACCAATATCTTAATAATAGACTCAATCTTCTTGTGTGTGCTGCTGCT 1199  
OY 1141 ACGAAAAGTTGGGGAAGATATGTAATTCGAAGAAAGGCAATCTCTGTTATGTCTTTGC 1200  
DB 1200 ACGAAAAGTTGGGGAAGATATGTAATTCGAAGAAAGGCAATCTCTGTTATGTCTTTGC 1259  
OY 1201 GAAAGATTTACATCTGAACCTGTTAAAAATCTTGAAGCAAGTTATCAAAAACAAGAG 1260  
DB 1260 GAAAGATTTACATCTGAACCTGTTAAAAATCTTGAAGCAAGTTATCAAAAACAAGAG 1319  
OY 1261 TGTTCACACACTTAATCTGCTAATAAAGAAAGTTCGCTCTGCGACCAAGAAATTTTA 1320  
DB 1320 TGTTCACACACTTAATCTGCTAATAAAGAAAGTTCGCTCTGCGACCAAGAAATTTTA 1379  
OY 1321 TGATTAAGCATATATCTGTTAATCTGAGGCTCATTAAGCCCTTGTGTTGMAAATAAGGCTG 1380  
DB 1380 TGATTAAGCATATATCTGTTAATCTGAGGCTCATTAAGCCCTTGTGTTGMAAATAAGGCTG 1439  
OY 1381 TAATTCGATTTCCAGCCTTACACAAATTAATTAAGAACGCTTGATGATGATGACTTAA 1440  
DB 1440 TAATTCGATTTCCAGCCTTACACAAATTAATTAAGAACGCTTGATGATGATGACTTAA 1499  
OY 1441 TAAAGAAAAATTTGATAGTATTAATGGAATTCCTGACCAATTCACCAATCCAGAGCG 1500  
DB 1500 TAAAGAAAAATTTGATAGTATTAATGGAATTCCTGACCAATTCACCAATCCAGAGCG 1559  
OY 1501 ACTTGGCAAAACCAATTCCTCAAAATTTAGATATCTGAAGAGCAAGTTGTTGCTCAATT 1560  
DB 1560 ACTTGGCAAAACCAATTCCTCAAAATTTAGATATCTGAAGAGCAAGTTGTTGCTCAATT 1619  
OY 1561 AGCTGATTAATATTAACAAGCTGATGTTTACATTTTGTGATGACATGATATTAATCACTGA 1620  
DB 1620 AGCTGATTAATATTAACAAGCTGATGTTTACATTTTGTGATGACATGATATTAATCACTGA 1679

```
QY 1621 TGAAGAGATGATATGTAAACGGCTCATATAGGCGCATAGTCACTGGATTGGAAAAGATAG 1680
    |||||||
Db 1680 TGAAGAGATGATATGTAAACGGCTCATATAGGCGCATAGTCACTGGATTGGAAAAGATAG 1739
QY 1681 CCTTTCGTGATAGAAAAAGTTCAGCTCAAGCCTATACATAAGAAAAAGGTATCCACC 1740
    |||||||
Db 1740 CCTTTCGTGATAGAAAAAGTTCAGCTCAAGCCTATACATAAGAAAAAGGTATCCACC 1799
QY 1741 TCCATCTCCGACGACGATGTTAAAGCAATCCAACTGGAGATAGTGCACACTATTTA 1800
    |||||||
Db 1800 TCCATCTCCGACGACGATGTTAAAGCAATCCAACTGGAGATAGTGCACACTATTTA 1859
QY 1801 CAATCGTGTAAAGGGGAAAAAGAAATCCACTGCTTCGACTTCATATATGGTTGAGCA 1860
    |||||||
Db 1860 CAATCGTGTAAAGGGGAAAAAGAAATCCACTGCTTCGACTTCATATATGGTTGAGCA 1919
QY 1861 TACAGTTGAGGTTAAAAACGGTAAATTTGATTTTCTCATAGGATCATTACATAATAT 1920
    |||||||
Db 1920 TACAGTTGAGGTTAAAAACGGTAAATTTGATTTTCTCATAGGATCATTACATAATAT 1979
QY 1921 TAAATTTGCTGGTTTGATGATACACATACAAAGCTCCAAATGGGTATCCCTGGAGA 1980
    |||||||
Db 1980 TAAATTTGCTGGTTTGATGATACACATACAAAGCTCCAAATGGGTATACCTGGAGA 2039
QY 1981 TTTGTTGCGACGATTAAAGTACTACGTAGAACACCCTGACGAACGTCCACATTTCTAATGA 2040
    |||||||
Db 2040 TTTGTTGCGACGATTAAAGTACTACGTAGAACACCCTGACGAACGTCCACATTTCTAATGA 2099
QY 2041 TGGATGGGGCAATGCCAGTAGAGATGTGTAGGCAAGAAAGACCACAGTAAATCCAAA 2100
    |||||||
Db 2100 TGGATGGGGCAATGCCAGTAGAGATGTGTAGGCAAGAAAGACCACAGTAAATCCAAA 2159
QY 2101 TAAGAACTTCAAAACCGGATGAAGAGCCAGTAGAGAAAACCTGCTGAGCCAGAACTCC 2160
    |||||||
Db 2160 TAAGAACTTCAAAACCGGATGAAGAGCCAGTAGAGAAAACCTGCTGAGCCAGAACTCC 2219
QY 2161 TCAGATGAGACGTGAAAAAGTAGAAGCCCAACTCAAAAGAGAGAAATTTGCTTGGCAA 2220
    |||||||
Db 2220 TCAGATGAGACGTGAAAAAGTAGAAGCCCAACTCAAAAGAGAGAAATTTGCTTGGCAA 2279
QY 2221 AGTAAGGATTTAGTCTGAAAGCCAAATGCAAGAAACCTAGCTGCTTACGAATAA 2280
    |||||||
Db 2280 AGTAAGGATTTAGTCTGAAAGCCAAATGCAAGAAACCTAGCTGCTTACGAATAA 2339
QY 2281 TTTGACTCTTCAATATATGATGAACAATAGTATCATGCGAGAGAGAAAAATTTACTTGC 2340
    |||||||
Db 2340 TTTGACTCTTCAATATATGATGAACAATAGTATCATGCGAGAGAGAAAAATTTACTTGC 2399
QY 2341 GTTGTAAAAAGAAATATCTTTCATCTGTAACTAAGGAAAAAATAAAC 2389
    |||||||
Db 2400 GTTGTAAAAAGAAATATCTTTCATCTGTAACTAAGGAAAAAATAAAC 2448

RESULT 4
AAVS2227
ID AAVS2227 standard; DNA; 8195 BP.
XX
AC AAVS2227;
XX
DE 23-OCT-1998 (first entry)
XX Streptococcus pneumoniae, genome fragment SEQ ID NO:94.
KW Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
computer readable medium; vaccine; pharmaceutical composition; ds.
XX Streptococcus pneumoniae.
OS
XX
XX PN MO9818931-A2.
XX 07-MAY-1998.
XX
```

```
PF 30-OCT-1997; 97WO-US19588.
XX
XX 31-OCT-1996; 96US-0029960.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PI Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M;
PI Kunsch CA, Rosen CA;
XX
DR WPI; 1998-272225/24.
XX
XX Computer-readable medium with recorded Streptococcus pneumoniae
PT polynucleotide sequences - useful in diagnostic kits and assays, and
PT pharmaceutical compositions and vaccines for Streptococcus
XX pneumoniae
PS Claim 1; Page 727-732; 1409pp; English.
XX
XX The present invention describes a computer readable medium which has
CC the nucleotide sequences SEQ ID NO:1 to 391 (AAV52134 to AAV52524)
CC recorded on it, or a representative fragment or a sequence at least 95%
CC identical to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in
CC SEQ ID NO:1 to 391 (AAV52134 to AAV52524) are genomic fragments from
CC Streptococcus pneumoniae. The present invention also describes an
CC isolated nucleic acid molecule encoding a homologue of any of the
CC fragments of the S.pneumoniae genome (SEQ ID NO:1 to 391) where the
CC nucleic acid molecule is produced by a process comprising: (a) screening
CC a genomic DNA library using as a probe a target sequence defined by any
CC of the sequences in SEQ ID NO:1 to 391, identifying members of the
CC library which contain sequences that hybridise to the target sequence and
CC isolating the nucleic acid molecules from the members; or (b) isolating
CC mRNA, DNA or cDNA produced from an organism, amplifying nucleic acid
CC molecules whose nucleotide sequence is homologous to amplification
CC primers derived from the fragment of the S. pneumoniae genome to prime
CC the amplification and isolating the amplified sequences. The computer
CC readable medium can be used in a computer-based system for identifying
CC fragments of the S. pneumoniae genome of commercial importance, or
CC expression modulating fragments of the S. pneumoniae genome. Products
CC from the present invention can be used in diagnosis kits and assays, and
CC pharmaceutical compositions and vaccines for S. pneumoniae.
XX
XX Sequence 8195 BP; 2688 A; 1622 C; 1777 G; 2105 T; 3 other:
XX
XX Query Match 100.0%; Score 2388; DB 19; Length 8195;
XX Best Local Similarity 100.0%; Pred. No. 0;
XX Matches 2388; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 TTTCTTACGAGTTGGGACTGTATCAACGCTTAAGAAAAATATCGTTTCCTA 60
    |||||||
Db 3053 TTCTTACGAGTTGGGACTGTATCAACGCTTAAGAAAAATATCGTTTCCTA 3112
QY 61 TATAGATGAAAAACAAGCGCGCAAAAAACGGAGATTTGACTCCGATGAGGTTAGCAA 120
    |||||||
Db 3113 TATAGATGAAAAACAAGCGCGCAAAAAACGGAGATTTGACTCCGATGAGGTTAGCAA 3172
QY 121 GCGTGAAGGAATCAATGCTGAGCAAAATCGTCATCAAGATACAGACCAAGCGTATGTCAC 180
    |||||||
Db 3173 GCGTGAAGGAATCAATGCTGAGCAAAATCGTCATCAAGATACAGACCAAGCGTATGTCAC 3232
QY 181 TTCACATGGCGACACTATCATTTATTACAATGGTAAAGTTCTTATGACGCTATCATCAG 240
    |||||||
Db 3233 TTCACATGGCGACACTATCATTTATTACAATGGTAAAGTTCTTATGACGCTATCATCAG 3292
QY 241 TGAAGAATTCATCAAGATCCAAACATTAAGCTAAAGAAAGTGAAGATGTTATATA 300
    |||||||
Db 3293 TGAAGAATTCATCAAGATCCAAACATTAAGCTAAAGAAAGTGAAGATGTTATATA 3352
QY 301 GGTCAAGGCTGATATGTATCAAGGTAGATGAAATATCTATGTTTACCTTAAGGATGC 360
    |||||||
Db 3353 GGTCAAGGCTGATATGTATCAAGGTAGATGAAATATCTATGTTTACCTTAAGGATGC 3412
QY 361 TGGCCACGGCGATTAACGTCCGTACAAAGAGAAATCAATCGACAAAAACAAGACATAG 420
    |||||||
```

Db 3413 TGCCACGGGATTAACGTCGTACAAAAAGAAATCAATCGACAAAAACAGAGCATAG 3472  
 QY 421 TCACATTCGTGAAGTGAACCTCCAGAAACAGATGGTGTGCGCTGGCAGCTTGCA 480  
 Db 3473 TCACATTCGTGAAGTGAACCTCCAGAAACAGATGGTGTGCGCTGGCAGCTTGCA 3532  
 QY 481 AGGAGCCTTACTACAGATGATGTTATATCTTTAATGCTTGTGATATCAGATAGATAC 540  
 Db 3533 AGGAGCCTTACTACAGATGATGTTATATCTTTAATGCTTGTGATATCAGATAGATAC 3592  
 QY 541 TGGTGTGCTTATATGCTTCTCATGAGATCATTTACATTACATTCCTTAAGATGATT 600  
 Db 3593 TGGTGTGCTTATATGCTTCTCATGAGATCATTTACATTACATTCCTTAAGATGATT 3652  
 QY 601 ATAGAGTGGAGTGGCTGCTGAGAACCTTCATCTGCTGCTGAGGAATGCTGAAA 660  
 Db 3653 ATAGAGTGGAGTGGCTGCTGAGAACCTTCATCTGCTGCTGAGGAATGCTGAAA 3712  
 QY 661 TTCAAGAACCTATGCGCGACAAAAATAGCGATTAACACTTCAAGAACAACTGGGTACCTTC 720  
 Db 3713 TTCAAGAACCTATGCGCGACAAAAATAGCGATTAACACTTCAAGAACAACTGGGTACCTTC 3772  
 QY 721 TGTAAAGCAATCCAGSAGTCAAACTACTACACAGAACAGACAGACACTAACAGTCA 780  
 Db 3773 TGTAAAGCAATCCAGSAGTCAAACTACTACACAGAACAGACAGACACTAACAGTCA 3832  
 QY 781 AGCAAGTCAAGTAAATGATGATGATGCTCTTGAAGACAGCTCTACAACTGGCTTTGAG 840  
 Db 3833 AGCAAGTCAAGTAAATGATGATGATGCTCTTGAAGACAGCTCTACAACTGGCTTTGAG 3892  
 QY 841 TCACAGCATGTAGAAATCTGATGCGCTGTCTTGTATCCAGCACAAATCCAAAGTCGAC 900  
 Db 3893 TCACAGCATGTAGAAATCTGATGCGCTGTCTTGTATCCAGCACAAATCCAAAGTCGAC 3952  
 QY 901 AGTAAAGGCTTGGCAGTGGCAGACAGAGATCATTTACACTTTCCTTACTCTCAAT 960  
 Db 3953 AGTAAAGGCTTGGCAGTGGCAGACAGAGATCATTTACACTTTCCTTACTCTCAAT 4012  
 QY 961 GTCTGAATTTGGAAGAGAAATCGCTGATTTATCCCTTGTATGCTTGAACCAATG 1020  
 Db 4013 GTCTGAATTTGGAAGAGAAATCGCTGATTTATCCCTTGTATGCTTGAACCAATG 4072  
 QY 1021 GGTACAGATTTCAAGGCCAGAACACCAAGTCCAGACCGAGACCTAGTCCAGG 1080  
 Db 4073 GGTACAGATTTCAAGGCCAGAACACCAAGTCCAGACCGAGACCTAGTCCAGG 4132  
 QY 1081 CCGGCAACCTGCGACCAAAATCTTAAATAGACTCAAAATCTTCTTGTGTTAGTACGCTG 1140  
 Db 4133 CCGGCAACCTGCGACCAAAATCTTAAATAGACTCAAAATCTTCTTGTGTTAGTACGCTG 4192  
 QY 1141 ACGAAAAAGTTGGGGAAGATATGATTCGAAGAAAAAGGAGCATCTCGTATGCTTTGCG 1200  
 Db 4193 ACGAAAAAGTTGGGGAAGATATGATTCGAAGAAAAAGGAGCATCTCGTATGCTTTGCG 4252  
 QY 1201 GAAAGATTTACATCTGAAAATCTGTTAAAAATCTTGAAGAGCAAGTTTCAAAACAGAG 1260  
 Db 4253 GAAAGATTTACATCTGAAAATCTGTTAAAAATCTTGAAGAGCAAGTTTCAAAACAGAG 4312  
 QY 1261 TGTTCACACACTTAACTGCTTAAAAAGAAATGTTGCTCGCTGAGCAAAATTTTA 1320  
 Db 4313 TGTTCACACACTTAACTGCTTAAAAAGAAATGTTGCTCGCTGAGCAAAATTTTA 4372  
 QY 1321 TGTATTAAGCATATATCTGTTAACTGAGGCTCATTAAGCCTTGTGTTGTTAAATTAAGGCTG 1380  
 Db 4373 TGTATTAAGCATATATCTGTTAACTGAGGCTCATTAAGCCTTGTGTTGTTAAATTAAGGCTG 4432  
 QY 1381 TAAATCTGATTTCCAAAGCCTTAGACAAATTTATGAAGCGTTGAATGATGATGACTAA 1440  
 Db 4433 TAAATCTGATTTCCAAAGCCTTAGACAAATTTATGAAGCGTTGAATGATGATGACTAA 4492  
 QY 1441 TAAAGAAAAATGCTAGATGATTTATGCTATCTAGACCAATTAACCATCCAGAGCG 1500  
 Db 4493 TAAAGAAAAATGCTAGATGATTTATGCTATCTAGACCAATTAACCATCCAGAGCG 4552

QY 1501 ACTTGGCAACCAAAATTTCTCAAAATGAGTATGATGAAGAGCAAGTTGATGCTCAAT 1560  
 Db 4553 ACTTGGCAACCAAAATTTCTCAAAATGAGTATGATGAAGAGCAAGTTGATGCTCAAT 4612  
 QY 1561 AGCTGATTAAGTATACAACTGATGATGTTTATGATGAATGATGATTAATCAAGTGA 1620  
 Db 4613 AGCTGATTAAGTATACAACTGATGATGTTTATGATGAATGATGATTAATCAAGTGA 4672  
 QY 1621 TGAAGAGATGCAATATGTAACGCTCATATGAGCCTATGCTACTGATTTGAAAAAGATAG 1680  
 Db 4673 TGAAGAGATGCAATATGTAACGCTCATATGAGCCTATGCTACTGATTTGAAAAAGATAG 4732  
 QY 1681 CATTGCTGATTAAGAAAAATTTGAGCTCAAGCTTACTTAAGAAAAAGATATCCAGC 1740  
 Db 4733 CATTGCTGATTAAGAAAAATTTGAGCTCAAGCTTACTTAAGAAAAAGATATCCAGC 4792  
 QY 1741 TCCATCTCCAGACGCAAGATGTTAAAGCAAAATCCAACTGAGATAGTCCAGCTATTTTA 1800  
 Db 4793 TCCATCTCCAGACGCAAGATGTTAAAGCAAAATCCAACTGAGATAGTCCAGCTATTTTA 4852  
 QY 1801 CAATCGTGAAGAGGGAAGAAAAAGAAATTCACATGCTTCCATATATGTTGAGCA 1860  
 Db 4853 CAATCGTGAAGAGGGAAGAAAAAGAAATTCACATGCTTCCATATATGTTGAGCA 4912  
 QY 1861 TACAGTTGAGTTAAAAACGTAATTTGATTTCTCATTAAGATCATTAACATAATAT 1920  
 Db 4913 TACAGTTGAGTTAAAAACGTAATTTGATTTCTCATTAAGATCATTAACATAATAT 4972  
 QY 1921 TAAATTTGCTTGTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1980  
 Db 4973 TAAATTTGCTTGTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATG 5032  
 QY 1981 TTTGTTTGGCAGCATTAAGTACTAGTACAGTACAGTACAGTACAGTACAGTACAGTAC 2040  
 Db 5033 TTTGTTTGGCAGCATTAAGTACTAGTACAGTACAGTACAGTACAGTACAGTACAGTAC 5092  
 QY 2041 TGGATGGGCAATGCGCAGTACAGTATGTTAGCAAGAAAGACACAGTGAATCCAAA 2100  
 Db 5093 TGGATGGGCAATGCGCAGTACAGTATGTTAGCAAGAAAGACACAGTGAATCCAAA 5152  
 QY 2101 TAAAGACTTCAAAAGCGGATGAAGAGCCAGTAGAGAAACCTGCTGAGCAGAGTCCC 2160  
 Db 5153 TAAAGACTTCAAAAGCGGATGAAGAGCCAGTAGAGAAACCTGCTGAGCAGAGTCCC 5212  
 QY 2161 TCAAGTAAAGACTGAAAAAATAGAACCCCACTCAAAAGAAAGCAAGTTTCTTGGCAA 2220  
 Db 5213 TCAAGTAAAGACTGAAAAAATAGAACCCCACTCAAAAGAAAGCAAGTTTCTTGGCAA 5272  
 QY 2221 AGTAAAGATTTAGTCTGAAAGCCAAATGCAAGAACTCTAGCTGTTAGGAATAA 2280  
 Db 5273 AGTAAAGATTTAGTCTGAAAGCCAAATGCAAGAACTCTAGCTGTTAGGAATAA 5332  
 QY 2281 TTTGACTCTTCAAAATTAAGATTAACAATAAGTATGATGCGAGAGCAAAAAATTAAGT 2340  
 Db 5333 TTTGACTCTTCAAAATTAAGATTAACAATAAGTATGATGCGAGAGCAAAAAATTAAGT 5392  
 QY 2341 GTTGTAAAAAGAGTAAATCTTCAATCTGTAAGTAAAGAAAAATAAC 2399  
 Db 5393 GTTGTAAAAAGAGTAAATCTTCAATCTGTAAGTAAAGAAAAATAAC 5441

RESULT 5  
 AAA65731  
 ID AAA65731 standard; DNB; 2523 BP.  
 AC AAA65731;  
 XX  
 XX  
 DT 21-NOV-2000 (first entry)  
 XX  
 DE Streptococcus pneumoniae BVH-11 gene SEQ ID NO:3.  
 XX  
 KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;

KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;  
KM otitis media; pneumonia; immunisation; bactericidal; ds.  
XX  
OS Streptococcus pneumoniae.  
XX  
PN WO200039299-A2.  
XX

PD 06-JUL-2000.

XX 20-DEC-1999; 99WO-CA01218.

PR 23-DEC-1998; 98US-0113800.

XX (BIOC-) BIOCHEM PHARMA INC.

PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;

DR WPI: 2000-452397/39.

XX P-PSDB: AAB12716.

XX Streptococcal antigens useful for vaccinating against e.g. meningitis,  
PT otitis media, bacteraemia and/or pneumonia -  
XX

PS Example 2; Fig 3; 106bp; English.

XX The present invention describes nucleic acids (I) encoding protein  
CC antigens (II) from Streptococcus pneumoniae. The protein antigens  
CC have bactericidal activity. The nucleic acids, encoding the protein  
CC antigens, may be used for the recombinant production of the proteins  
CC they encode. The protein antigens may then be used as vaccines for the  
CC prevention and treatment of Streptococcal infections in mammals  
CC (especially humans) which result in, e.g. meningitis, otitis media,  
CC bacteraemia and/or pneumonia. The present sequence encodes the  
CC S. pneumoniae BVH-11 protein antigen.

XX Sequence 2523 BP; 879 A; 523 C; 526 G; 595 T; 0 other:

Query Match 57.5%; Score 1374.2; DB 21; Length 2523;

Best Local Similarity 73.9%; Pred. No. 0;

Matches 1819; Conservative 0; Mismatches 539; Indels 102; Gaps 2;

QY 1 TTTCTTCAGATGGGACTGTATCACTAGACGGTTAAGAAAATATTCGTTCCTA 60  
DB TGGTTATGAACTAGTTCATCAAGCTCAAACTGTPAAAGAAATATCGTGTTCCTA 119  
QY 61 TATAGATGAAAAACAAGCGACGCAAAAAAGAGAAATTTGACTCTGTAGAGTTAGCAA 120  
DB TATAGATGAAAAACAAGCGACGCAAAAAAGAGAAATTTGACTCTGTAGAGTTAGCAA 179  
QY 121 GCGTGAAGGATCAATGCTGAGCAAAATCGTCATCAAGATACAGACCAGGCTATGTCAC 180  
DB GCGTGAAGGATCAATCAAGCGGACGCAAAATCGTCATCAAGATACAGGATATGTCAG 239  
QY 181 TTTCACTGGGACCACTATCATTTATTAAGTGAAGTTCCTTATGAGGCTTATCAAC 240  
DB CTCTCATGGAGACCATTAATCATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 299  
QY 241 TGAAGATTAATCAATGAAGATCAAACTAATGAAGTAAAGATGAGGATATGTTATATGA 300  
DB TGAAGATTAATCAATGAAGATCAAACTAATGAAGTAAAGATGAGGATATGTTATATGA 359  
QY 301 GGTCAAGGCTGATATGTTATCAAGTAGAGAAAATTAATTAATTAATTAATTAATTAAT 360  
DB AATCAAGGCTGATATGTTATCAAGTAGAGAAAATTAATTAATTAATTAATTAATTAAT 419  
QY 361 TGGCCAGGCGGATTAAGTCCGTCAAAAGAGAAATCAATCAAGCAAAACAAAGAGATG 420  
DB AGTTCATGGGATTAAGTCCGTCAAAAGAGAAATCAATCAAGCAAAACAAAGAGATG 479  
QY 421 TCAACATCGGAGGATGAGAACTCAAGAAAAGATGCTGCTGCTGGCAGCTTGCAG 480  
DB TCAACATCGGAGGATGAGAACTCAAGAAAAGATGCTGCTGCTGGCAGCTTGCAG 539

QY 481 AGAGCGTATACTACAGATGATGTTATATCTTTAATGCTTGTATATCATAGAGATAC 540  
DB GAGACGCTACACAGATGATGTTATATCTTTAATGCTTGTATATCATAGAGATAC 599  
QY 541 TGGTATGCTTATATGCTTCCATGAGATGATTAATCAATTAATTAATTAATTAATTAAT 600  
DB GAGCGATGCTTATATGCTTCCATGAGATGATTAATCAATTAATTAATTAATTAATTAAT 659  
QY 601 ATCAGCTACGAGTTGGCTGCTGAGAAAGCTTCCATTCGTGAGAAATCTGTCAA 660  
DB ATCAGCTACGAGTTGGCTGCTGAGAAAGCTTCCATTCGTGAGAAATCTGTCAA 719  
QY 661 TTCAGAACCTTATCGCCGCAAAATATGCGATTAACCTTAAGAACAACTGGTACTTC 720  
DB TTTAAGAACCTTATCGCCGCAAAATATGCGATTAACCTTAAGAACAACTGGTACTTC 779  
QY 721 TGTAAAGCAATCCAGAACATCAAAATTAATCAACCAAGCAACAGCACTAACAGTCA 780  
DB TGTAAAGCAATCCAGAACATCAAAATTAATCAACCAAGCAACAGCACTAACAGTCA 839  
QY 781 AGCAAGTCAAAATTAATGATTAATGATGATGCTCTTGAAGACGCTCTACAACTGCTTTGAG 840  
DB AGCAAGTCAAAATTAATGATTAATGATGATGCTCTTGAAGACGCTCTACAACTGCTTTGAG 899  
QY 841 TCAAGCAATGTAATGATGATGCTCTTGAAGACGCTCTACAACTGCTTTGAG 900  
DB TCAAGCAATGTAATGATGATGCTCTTGAAGACGCTCTACAACTGCTTTGAG 959  
QY 901 AGCTAAGGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960  
DB AGCTAAGGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1019  
QY 961 GTCGTATGGAAGAAAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020  
DB GTCGTATGGAAGAAAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1079  
QY 1021 GGTACCAATTTCAAGGCGCAACCAACCAAGTCCACACGAGTCCGGAACCTATGTCAG 1080  
DB GGTACCAATTTCAAGGCGCAACCAACCAAGTCCACACGAGTCCGGAACCTATGTCAG 1139  
QY 1081 CCGGCAACCTGCAACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1128  
DB CCGGCAACCTGCAACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1199  
QY 1129 TATGAGCTGTAAGAAAGTTGGGAGAGATATGATTTGAAGAAAAGGATCTCTCG 1188  
DB TATGAGCTGTAAGAAAGTTGGGAGAGATATGATTTGAAGAAAAGGATCTCTCG 1259  
QY 1189 TATGCTTTGCAAGATTTTACATCTGAACCTGTTAAATTAATTAATTAATTAATTAAT 1248  
DB TATGCTTTGCAAGATTTTACATCTGAACCTGTTAAATTAATTAATTAATTAATTAATTAAT 1319  
QY 1249 AAACCAAGAGAGTGTTCACACACTTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1308  
DB AAACCAAGAGAGTGTTCACACACTTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1379  
QY 1320 CAAGCAAGAAAGTTTATCTCATATAGCTAGAGCTAAGAAATTAATTAATTAATTAAT 1379  
DB CAAGCAAGAAAGTTTATCTCATATAGCTAGAGCTAAGAAATTAATTAATTAATTAATTAAT 1439  
QY 1309 CCAAGAAATTTATGATTAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1368  
DB CCAAGAAATTTATGATTAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1439  
QY 1429 TGAATGACTTAATTAAGAAATTTGTTAATGATTAATTAATTAATTAATTAATTAATTAAT 1488  
DB TGAATGACTTAATTAAGAAATTTGTTAATGATTAATTAATTAATTAATTAATTAATTAATTAAT 1559  
QY 1489 CCATTCAGAGCACTTGGCAAACTCAAAATTTGAGTATTAATTAATTAATTAATTAATTAAT 1548  
DB CCATTCAGAGCACTTGGCAAACTCAAAATTTGAGTATTAATTAATTAATTAATTAATTAATTAAT 1619  
QY 1549 TATTCCTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1608

```

Db      1620 AGTAGCCAACTTGGCAGGCAAGTACACACAGAGAGCGTTATCTTTATCCCTCGTA 1679
QY      1609 TATTAATGAGTGAAGAGATGATGTAAAGCCCTCATATGGCGATGCTCGAT 1668
Db      1680 TATTAACCAAGTATGAGGGGATGCTATGTAACTCCATATATGACCATGCGATGAT 1739
QY      1669 TGGAAAAAGATAGCCTTCTCTATAGAAAAAGTTGACGCTCAAGCCCTATCTAAAGAAA 1728
Db      1740 TAAAAAAGATAGTTTGTCTAGAGTGAGAGCGCGACGCCAGCTTATGCTAAAGAA 1799
QY      1729 AGGTATCTTACCTTCATCTCAGAGCAGATGTTAAAGCAATCCAACTGAGATAGTC 1788
Db      1800 AGGTTTACCCCTCCCTCGACAGCCATCAGGATTCAGGAATATCTGAGCAAAAAGGAGC 1859
QY      1789 AGCAGCTATTTAACATGATGTTGAAGGGGAAAAAGCAATCCCTGCTGAGCTCCATA 1848
Db      1860 AGAAGCTATTTACACCGCTGAAAGCAGCTAAGAGAGCTGCACTTATGATGCTTA 1919
QY      1849 TATGTTGACATACAGTTGAGTTAAAAACGGTATTTGATTATTCCTATAAGATCA 1908
Db      1920 CAATCTTCAATATACGTGTAGAGTCAAAAACGGTATTTATCATACCTATATATGACCA 1979
QY      1909 TTTACATTAATTAATTTCTGTTGTTGATGATCACACATPACAAAGCTCCAAATGGCTA 1968
Db      1980 TTTACATTAATTAATTTCTGTTGTTGATGATCACACATPACAAAGCTCCAAATGGCTA 2039
QY      1969 TACCTTGAAGATTTGTTTGGCAGATTAAGTACTAGTGAACACCCCTACGAACTCC 2028
Db      2040 TACTTGAAGATTTGTTTGGCAGTGTCAAGTACTAGTGAACATCAACGAAAGCTCC 2099
QY      2029 ACATTTATATGATGATGGGCAATGCCAGTGAAGATGTTAGGCAAGAAGCCACAG 2088
Db      2100 GCAATTCAGATTAATGGTTTGTAAAGCTAGCGACCATGTTCAAGAAACAAANAATGCTCA 2159
QY      2089 TGAAGATCCAAATPAGAACTTCAAGCGGATGAAGAG----- 2125
Db      2160 AGCTGATACCAATCAAAACGGAACCAACGAGAGAAACCTCAGACAGAAAAACCTGA 2219
QY      2126 ----- 2125
Db      2220 GGAAGAAACCCCTCGAGAGAGAAACCAACAAAGCGAANAACAGATCTCCAAAAACCAAC 2279
QY      2126 -----CCAGTAGAGAGAAACACCTGCTGAGCGAAGATCCCTCAATAGAGACTGAAAA 2178
Db      2280 AGAGGAACCGAAGAAAGAAATCACCAGAGAAATCAGAAAGAACTCAGTCCAGACTGAAAA 2339
QY      2179 AGTAGAAGCCCAACTCAAGAGAGAGAGTTTGGCTTGGCAAGTAAGGATCTTACTCT 2238
Db      2340 GGTGGAAGAAAAAACTGAGAGAGGCTGAAGATTACTTGAATAAATCCAGATCCAAATTAT 2399
QY      2239 GAAGGCAATGCAACAGAACTAGCTGTTTACGAAATAATTTGACTCTTCAAAATTAT 2258
Db      2400 CAAGTCAATAGCCAAAGAGAGCTCTCAGAGATTAAAAAATTAATTTACTATTATTTGGACCCA 2459
QY      2299 GGATTAACAATAGTATCATGCGAGAGAGAGAAAAATTTACTTGGTGTGTTAAAGGAAGTAA 2358
Db      2460 GGACAAACAATACATATATATGCGAGAGCTGAAAAAATTTGGCTTTATTTAAAGAGAGTAA 2519

```

```

RESULT 6
ID      AAA65736
ID      AAA65736 standard; DNA; 2647 BP.
XX
XX      AAA65736.
XX
XX      21-NOV-2000 (first entry)
DE      Streptococcus pneumoniae BVH-11 gene SEQ ID NO:12.
XX
XX      Streptococcus pneumoniae, BVH-3; BVH-11; BVH-28; antigen; vaccine;
KM      Propylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
KW      Otitis media; pneumonia; immunisation; bactericidal; ds.

```

```

XX      OS      Streptococcus pneumoniae.
XX
XX      MO200039299-A2.
XX
XX      PD      06-JUL-2000.
XX
XX      PF      20-DEC-1999; 99MO-CA01218.
XX
XX      PR      23-DEC-1998; 98US-0113800.
XX
XX      PA      (BIOC-) BIOCHEM PHARMA INC.
PI      Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
DR      WPI; 2000-452397/39.
XX
XX      PT      Streptococcal antigens useful for vaccinating against e.g. meningitis,
XX      PT      otitis media, bacteraemia and/or pneumonia -
XX      PS      Example 6; Fig 15; 106pp; English.
XX
XX      CC      The present invention describes nucleic acids (I) encoding protein
XX      CC      antigens (II) from Streptococcus pneumoniae. The protein antigens
XX      CC      have bactericidal activity. The nucleic acids, encoding the protein
XX      CC      antigens, may be used for the recombinant production of the proteins
XX      CC      they encode. The protein antigens may then be used as vaccines for the
XX      CC      prevention and treatment of Streptococcal infections in mammals
XX      CC      (especially humans) which result in, e.g. meningitis, otitis media,
XX      CC      bacteraemia and/or pneumonia. The present sequence encodes the
XX      CC      S. pneumoniae BVH-11 protein antigen.
SQ
SQ      Sequence 2647 BP; 934 A; 538 C; 556 G; 619 T; 0 other;
Query Match 57.5%; Score 1374.2; DB 21; Length 2647;
Best Local Similarity 73.9%; Pred. No. 0;
Matches 1819; Conservative 0; Mismatches 539; Indels 102; Gaps 2;
QY      1 TTTCTAGAGTTGGGACTGTATCAAGCTAGACGCTTAAGGAAATTAATGCTTCTCTA 60
Db      104 TGTCTATGAACCTAGGTTTGATCAAGCTCAAACTGTAAGAAATAATGCTGTTCTTA 163
QY      61 TATAGATGAAAAACAAACGACGCAAAAAACGAGAAATTTGACTCTGATAGTTAGCAA 120
Db      164 TATATAGTGAAGAAACAAACGACGCAAAAAACGAGAAATTTGACTCTGATAGTTAGCAA 223
QY      121 GCGTGAAGAAATCAATCTGAGCAAAATGCTCATCAAGATTAACAGACCAAGGCTATGTCAC 180
Db      224 GCGTGAAGAAATCAACGCCGCAAAATGCTCATCAAGATTAACAGATTAACAGGTTATGTGAC 283
QY      181 TTGCATAGGCGGACCACTATGATTAATTTCAATGTAAGGCTCCTATACGCTATACACAG 240
Db      284 CTCTCATGAGAGACCATTAATATCTAATTAATGTAAGGCTCCTATATGATCCATCATCAG 343
QY      241 TGAAGATTTACTCTCAAGAAATCAAACTTAAGCTTAAGCTTAAGGATTAATTTGTTATGA 300
Db      344 TGAAGACTCTCTCATGAAGATTCGAATTTATCACTTGAAGATTTACAGACTTGCATATGA 403
QY      301 GGTCAAGGTTGATGTTATCAAGGTAGATGGAATAATTAATGTTTACCTTAAGATGC 360
Db      404 AATCAAGGTTGTTATCTAATTAAGTAAACGGTAATTAATTAATGTTTACCTTAAGATGC 463
QY      361 TGGCCAGCGGATTAATCTCGGTACAAAGAGGAATTAATTCAGACAAAAACAAGGCTAG 420
Db      464 AGCTCATGCGGATTAATCTCGGTACAAAGAGGAATTAATTCAGACAAAAACAAGGCTAG 523
QY      421 TCAACATCTGTAAGGAGTGAAGTCAAGAAAGATGAGTGTGTTGCTTGGCAGCTTGGCA 480
Db      524 TCAAGCATCGTAAGAGAGGAGCTTCAAGAAAGATGAGTGTGTTGCTTGGCAGCTTGGCA 583
QY      481 AGGACGCTATACATACAGATGATGTTATTAATCTTGAATGCTTGTGATATCATAGAGATAC 540
Db      584 GGGACGCTACACACAGATGATGTTATTAATCTTCAATGATCATGATCATGCAAGATAC 643

```

[illegible]

QY	1609	TATATACTGATATAGAGAGATGCTATATGTAAAGCCTCATATGGCCATATGACTGGAT	1668
Db	1724	TATATACAGTATAGAGGGAGTGCCTATGTAACTCCATATGATGACCATAGCCATGGAT	1783
QY	1669	TGGAAACATATGACCTTTCTGTATTAAGAAAAAGTGTGAGCTCAAGCCTATACTAAAGAAA	1728
Db	1784	TAAAAAGATATGTTTGTCTGAGCTGAGAGAGCGGACGCCAGCCTTATGTTAAAGAGA	1833
QY	1729	AGGTATCTTACCTCCATCTCCAGACGCAGATGTTAAAGCAAAATCCAACTGAGATATGTC	1788
Db	1844	AGGTTTGACCCCTCTCCAGACGCATCGATTCGAGAAATTAAGAAATTAAGAGAA	1903
QY	1789	AGCAGCTATTTACATATGCTGTGAAAGGGGAAAAAGCAATTCACCTGTTCCACTTCATA	1848
Db	1904	AGAAAGCTATCTACACCGCGGTGAAGAGCTAAAGAAAGGTGCCATGTGATGCTCTTA	1963
QY	1849	TATGTTAGAGATACACTGTAGAGGTTAAAAAGGTAATTTGATTTATTTCCATTAAGATCA	1908
Db	1964	CAATCTTCAATATCTGTATAGACCAAAAAGGTATTTAACTATACCTCATATAGACCA	2023
QY	1909	TTACCATTAATTTAAATTTGCTGTTTATGATCATACATACAAAGCTCCAAATAGCTA	1968
Db	2024	TTACCATTAACATCAAAATTTGAGTGGTTTGACGAAGCCCTTTATGAGGACCTTAAGGGTGA	2083
QY	1969	TACCTTGGAGATTTGTTTGGCAGCATTAATATCTACGTAGAACACCTGACGAACCTCC	2028
Db	2084	TACTCTTGAGATCTTTTGGCGAGCTGCAAGTACTATGTGCAAGATCCAAAGCAACCTCC	2143
QY	2029	ACATTTCAATGATGATGGGGCAATGCGATGCGATGCTTTAGGCAAGAAAGCCACAG	2088
Db	2144	GCATTCAGATATAGTGTGTTGGTAACGCTAGCGCACTGTTCAAAGAAACAAAATGTCTCA	2203
QY	2089	TGAAGATCCAAATAGAACTTTCCAAGCGATGAAGAG-----	2125
Db	2204	AGCTGATATCCATATCAAGGAAAAACCAAGGAGGAGAAACCTCAGACAGAAAAACCTGA	2263
QY	2126	-----	2125
Db	2264	GGAAGAAACCCCTCGAAGAAAGAAACCAAGGCGAAGAACCGAGTCTCCAAAAACCAAC	2323
QY	2126	-----CCAGTAAAGAAACACCTGCTGAGCCAGAAAGTCCCTCAAGTAAAGAGACTGAAAA	2178
Db	2324	AGAGGAACCCAGAAAGAAATACACCAAGAGAAATGAGAAACCTCAGCTGAGACTGAAAA	2383
QY	2129	AGTAAAGCCCAATCTCAAAAGACAGAAAGTTTGCCTTGGGAAGTAAACGATTTAGTCT	2238
Db	2384	GCTTGAAGAAAAACTGAGAGAGCTGAAATTTACTTGGAAAAATCCAGATCCAAATAT	2443
QY	2239	GAAAGCCATGCAACAGAAACCTAGCTGGTTTACGAAATAATTTGACTCTTCAATTTAT	2298
Db	2444	CAAGTCCAAATGCCAAAGAGACTCTCACAGAGATTTAAAAAATTAATTTACTATTTGGCACCA	2503
QY	2239	GGATTAACATATGATTCATGGCAGAAACAGAAAAATTTACTGGCTTGTAAAAAGAGTAA	2358
Db	2504	GGACACAAATATATATATATGCGCAAGACCTGAAAAAACTATTTGCTTATTTAAAGAGATTA	2563
RESULT 7			
ABK15103			
ID	ABK15103 standard; DNA; 2647 BP.		
XX	ABK15103;		
XX	AC		
XX	DT		
XX	08-MAY-2002 (first entry)		
DE	DNA encoding Streptococcus pneumoniae BVH-11.		
XX			
KW	BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;		
XX	pneumonia; streptococcal bacterial infection; gene; ds.		
OS	Streptococcus pneumoniae.		
XX			
Key	Location/Qualifiers		
PH			





Db 1424 TCGAGAAATTTTACAAATAGGCTTATGCTTACTACTAGCAGAAATTCACCAAGATTACTTGA 1483  
QY 1369 AAATPAAGGTCGTAAATCTGATTTTCCAAAGCCTTAGACAAATTTATTAGACGCTTGAATGA 1428  
Db 1484 TAATPAAGGTCGACAAATGTTGATTTTGAGCTTGTGATTAACCTGTTGAGACGACTCAAGGA 1543  
QY 1429 TGAATCGACTTAATTAAGAAAAATTTGGTAGATGATTATTGGCATTTCTAGCACCATTAC 1488  
Db 1544 TGTCTCAAGGATTAAGTCAAGTAGTGATGATATTCTGCTCCCTTACTACTCCGATTCG 1603  
QY 1489 CCATCCAGAGCCTTGGCAACCAATTTCTCAATATGATATACATGTAAGCAAGATTCG 1548  
Db 1604 TCATCCGAAGCCTTAGGAAAAACCAATTCGCAATTTACCTACTACTATGATGATTTCA 1663  
QY 1549 TATTGCTCAATTAAGCTGATAGATATACACAGCTGATGTTTACTATTTTGTATGAACATGA 1608  
Db 1664 AGTAGCCCAAGTTGGCAGCAAGTACACAGAAAGAGGTTTATATCTTTGATCTCTGTGA 1723  
QY 1609 TATATATAGATGAGAGAGATGATATGTAAGCCTCATATGCGCATATGTCATGAT 1668  
Db 1724 TATATACCAGTATGAGGCGATGCTATGTACTCCACATATGACCATAGCCACTGAT 1783  
QY 1669 TGGAAAGATAGCCTTCTGATTAAGAAAAAGTTGCAGCTCAAGCCTATCTATAAGAAA 1728  
Db 1784 TAAAAAGATAGTTTGTCTGAGCTGAGAGAGGCGCCAGCCTTATGCTTAAAGGAA 1843  
QY 1729 AGGTATCCATCTCATCTCCAGACGAGATGTTAAAGCAATCCAACTGAGATAGTGC 1788  
Db 1844 AGGTGAGCCCTTCCTTCGACAGCCATCAGATTCAGAAATCTAGAGCAAAAGAGAGC 1903  
QY 1789 AGCAGCTATTTAACATTCGTGAAAGGGGAAAAACGAATTCACCTGCTTGACTTCATA 1848  
Db 1904 AGAAGCTATCTACACGCGCTGAAAGAGAGCTAAGAAAGCTGATGCTATGCTTGA 1963  
QY 1849 TATGTTGACATACAGTTAGTTTAAACGGTAATTTGATTTATTCCTATAGAGATCA 1908  
Db 1964 CATCTTCAATATCTGTAGAAAGTCAAAAAAGCTAGTTTATATATCTATATATGACCA 2023  
QY 1909 TTACATATATTTAAATTTGCTTGTGATGATCAACATACAAAAGCTTCAAAATGGCTA 1968  
Db 2024 TTACATATATCAAAATTTGAGTGTGACGAAGGCGCTTATGAGCACCCTTAAGGGTGA 2083  
QY 1969 TACCTTGAAGATTTTGTTCGACGATTAAGTACTAGTGAACACCCCTTACAGAGCTCC 2028  
Db 2084 TACTCTGAGATCTTTTGGCGACTGTCAGTACTATGTCGAACATCCAAACCAAGCTCC 2143  
QY 2029 ACATTTAATGATGATGGGCAATGCCAGTGAAGCATGTTTGAAGCAAGAACACACAG 2088  
Db 2144 GCATTGAGTAAATGCTTTTGTGTAACGCTAGCCATGTTCAAAAGAAACAAAAATGTCTA 2203  
QY 2089 TGAAGATCCAAATPAAGAACTTCAAAGCGGATGAAGAG----- 2125  
Db 2204 AGCTGATACCAATCAAAACGGAACCAAGCGAGAGAAACCTCAGACAGAAACACCTGA 2263  
QY 2126 ----- 2125  
Db 2264 GGAAGAAACCCCTCGAAGAGAGAAACCAAAAGCGAAGAACAGAGTCTCCAAACCAAC 2323  
QY 2126 -----CCAGTAGAGAAACACCTGCTGAGCCAGAAAGTCCCTCAAGTAGAGACTGAAA 2178  
Db 2324 AGAGGAACCAAGAAAGAAATACACAGAGAAATCAGAAAGAACTCAGAGCTGAGCTGAAA 2383  
QY 2179 AGTAGAAGCCCAACTCAAAAGAACAGAGTTTGTTCGGAAGTAACGATTTCTAGTCT 2238  
Db 2384 GGTGAAGAAAAAAGTGAAGAGCTGAAGATTTACTTGGAAAAATCCAGATCCAAATTAAT 2443  
QY 2239 GAAAGCAATGCAACGAAGAACTTACTGTTTACGAATTAATTGACCTCTTCAATTTAT 2298  
Db 2444 CAAGTCAATGCGCAAGAGACTCTTCACAGGATTTAAAAAATTAATTTCTATTGGCCACCA 2503  
QY 2299 GGATTAACATAGTATCATGGCAGACAGAAAAATTAATTTACTGCTTTTAAAGAGATTA 2358  
Db 2504 GGACAAATATCTATTTATGCGAGAGAGCTGAAAAACCTATTGGCTTTTATTAAAGAGAGTAA 2563

RESULT 8  
ID AAA08557  
XN AAA08557 standard; DNA; 2478 BP.  
XN AAA08557;  
AC 19-JUL-2000 (first entry)  
DT 19-JUL-2000 (first entry)  
XN  
XN  
DE S. pneumoniae 92 kDa human C3-degrading protein coding sequence.  
XN  
XN Human C3-degrading protein; 92 kDa; immunostimulatory; vaccine;  
XN Inhibitor; inflammation; organ rejection; xenotransplantation; ss.  
OS Streptococcus pneumoniae.  
PN MO200017370-A1.  
PD 30-MAR-2000.  
PF 24-SEP-1999; 99WO-US22362.  
PR 24-SEP-1998; 98US-0101736.  
PR 31-MAR-1999; 99US-0283094.  
XN  
XN (MING ) UNIV MINNESOTA.  
XN (AMCY ) AMERICAN CYANAMID CO.  
PI Hostetter MK, Finkel DJ, Cheng Q, Green BA, Masi AW;  
XN  
XN WPI: 2000-283594/24.  
DR P-PSDB; AAY91939.  
PT Isolated polypeptide is used to stimulate immune system and immunize or  
PT treat a mammalian subject against Streptococcus pneumoniae infection or  
PT colonization  
PS Claim 1, Page 55-57; 63pp; English.  
XX  
XX The present sequence, isolated from Streptococcus pneumoniae, encodes a  
CC human C3-degrading protein (see AAY91939) of about 92 kDa. This  
CC sequence may encompass a smaller 20 kDa polypeptide coding sequence  
CC (AA08556) also having human C3-degrading activity.  
CC The DNA sequences can be used for producing an immune response to  
CC Streptococcus pneumoniae in a mammal. Antibodies against the proteins  
CC can be used to inhibit S. pneumoniae-mediated C3 degradation.  
CC C3-mediated inflammation and rejection in xenotransplantation can be  
CC inhibited by expressing the nucleic acid sequences on the surface of an  
CC organ of an animal. In particular, the polypeptides are useful for  
CC stimulating the immune system and are effective to immunize or treat a  
CC mammalian subject against Streptococcus pneumoniae infection or  
CC colonization.  
XX  
SQ Sequence 2478 BP; 837 A; 510 C; 535 G; 596 T; 0 other:  
Query Match 42.4%; Score 1011.8; DB 21; Length 2478;  
Best Local Similarity 66.2%; Pred. No. 3.8e-244;  
Matches 1628; Conservative 0; Mismatches 688; Indels 144; Gaps 6;  
QY 1 TTTCTTCGAGTTGGGAGCTGTATCAAGCTGAACGGTTAAGAAAA--TAATCTGTGTTTC 57  
Db 60 TTCTCTATGAACTTGGTCGTCACCAAGCTGTCAAGTTAAGAAAGAGTCAATTCGAGTTTC 119  
QY 58 CTATATAGATGGAAGAAACGAGCGCAAAAGCGGAATTTGACTCCGATGAGAGTTAG 117  
Db 120 TTATATAGATGAGTCAAGCTGTGTCAAAAGCAGAAAACCTTGACCCGAGATGAAGTCA 179  
QY 118 CAAGCGTGAAGAAATCAAGTGTGACCAATTCGTCAATCAAGATTAAGAACCAAGCGTATGT 177  
Db 180 TAAAGAGGAGGGGATCAAGCGCGGAACAAATTCGTCAATCAAGATTAAGGATCAAGGTTATCT 239  
QY 178 CACTTCACATGCGCAGCACTATCATTAATTAACAATGGTAAGTCTTATGACGCTATCAT 237

Db 240 GACCTCATGAGACCATATCTACTATATGCGAAGCTCCCTATATGCAATCAT 299  
Qy 238 CAGTGAAGATTTACTCATGGAAGATCCAACTTAAGCTAAAGATGAGATATGTTAA 297  
Db 300 CAGTGAAGACCTCTCTGGAAGATCCGAATTTATCACTTAAGAGATTCACACATTTGCAA 359  
Qy 298 TGAGTCAAGGGTGAATGTATTCAGAGTAGATGGAATAATCTATGTATACCTTAAGGA 357  
Db 360 TGAATCAAGGGTGTATTTATTCAGAGTAGATGGAATAATCTATGTATACCTTAAGGA 419  
Qy 358 TCGTCCCAAGCGGATACCTCCGTACAAAAGAGAAATCAATCGACAAAACAGAGCA 417  
Db 420 TCGAGCTCATGCGGATATATTCGCAAAAAGAGATTTAAAGCTGAGAAGCAGGAACA 479  
Qy 418 TACTCAACATCGTGAAGGTGAACTCCAAAGACGATGGGTGCTCCCTGGACGCTTC 477  
Db 480 CAGTCAATATCAGGGGGGTGTCT-----AACGATCAAGCAGATGTTGCAAGCCAGAGC 533  
Qy 478 GCAAGGACGCTATACTACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGA 537  
Db 534 CCAAGGACGCTATACAAAGGATGTTATCTTCAATGCAATCTGATATCATGTAGGA 593  
Qy 538 TACTGATGCTTATATCTCTCTCTCAATGAGATCATTTACCATTTACCTCTTAAGATGA 597  
Db 594 CACGGGTGATGCTTATATCTCTCTCAAGCGACCATTTACCATTTACCTCTTAAGATGA 653  
Qy 598 GTTATACGATGAGGATGTGGTGTGTCAGAAAGCTTCTCTATCTGCTGAGGAATCTGTC 657  
Db 654 GTTATACGATGAGGATGTGGTGTGTCAGAAAGCTTATGGAATGG----- 698  
Qy 658 AATTCGAAGACCTATCGCCGACAAAATAGCGATACACTTCAAGAACAACTGGGTACC 717  
Db 699 -GAAGACGAGATCTCGTCTCTTCAAGTCTGTATATGCAAAATCCAGCTCAACCAA 757  
Qy 718 TTCTGTACGAATCCAGGAACCTACAAATACTAACACAGAACACAACTAAACAG 777  
Db 758 GATTGTGAGGAACCAACATCT-----GACTGTCACTCCAACTTA 797  
Qy 778 TCAAGCAAGTCAAGATGATGATGTAGTCTTGAAGAGCTGTACAAAGCTGCTT 837  
Db 798 TCAATCAAAATCAAGGGAACCAATTTCAAGCCTTTTACGTAATGTATGCTTAACCTT 857  
Qy 838 GAGTCAACGATGTAGATCTGATGCTGTGTCTTTGATCCAGACAAATTCACAACTGC 897  
Db 858 ATCGAAGCGCATGTGGAATCTGATGCTTATTTGCAACCGAAGCAAAATCAAACTGC 917  
Qy 898 AACAGCTAGAGGTGTGCAATGTCACACGAGATCATTTACCTTCACTCCCTTACTCTCA 957  
Db 918 AACCGCAGAGGTGTACTCTCCCTCATGTATACCATTTATCCCTTATGAGACA 977  
Qy 958 AATGTGATTTGGAAGAGAAATCGCTGATATTTCCCTTGATGCTTAAACCA 1017  
Db 978 AATGTGATTTGGAAGAGAAATCGCTGATATTTCCCTTGATGCTTAAACCA 1037  
Qy 1018 TTGGGTACCGAGATTCAGAGCCAGAACCAAGTCCAAACGCACTCCGAACTAGTCC 1077  
Db 1038 TTGGGTACCGAGATTCAGAGCCAGAACCAAGTCCAAACGCACTCCGAACTAGTCC 1097  
Qy 1078 AGGCCCGCAACCTGACCAATCTTAAATAGACTCA-----ATTCTTCTTT 1125  
Db 1098 AAGTCCGCAACCTGACCAATCTTAAACAGCTCCAAAGCAATCCAAATTTGAGAAATT 1157  
Qy 1126 GGTATGACTGCTGTGGAAGATGTGGGAAGATATTTGCGAAGAAAAGGCAATCTC 1185  
Db 1158 GGTCAAGAAAGCTGTGGAAGATGTGGGAAGATGTGGGAAGATGTGGGAAGATGTGG 1217  
Qy 1186 TCGTATATGCTTTCGGAAGATTTACCATCTGAAAGCTTTAAAAATTTGAAAGCAATT 1245  
Db 1218 TCGTATATATCCAGCCAAAGATCTTTCAGAGAAACAGACGAGCATGTATACCAACT 1277  
Qy 1246 ATCAAAACAAAGAGAGTCTTTCACACACTTTTAACTGCTTAAAAAGAAATGTGCTCTCG 1305

Db 1278 GCGCCAGACGAAAAGTTTATCTCATTAAGCTAGAGACTAAGAAAATGACCTCCATCTAG 1337  
Qy 1306 TGACCAAGAAATTTATGATAAAGCAATATCTGTATAGTAGGCTCATAAAGCTTGT 1365  
Db 1338 TGATCGAAGATTTTACATTAAGCTTATGACTTACTGCAAGAAATTCACCAAGATTACT 1397  
Qy 1366 TGNAAATGAAGGTCTGAATTTCTGATTTCCAAAGCTTTGAGCAAAATTTATAGAACGCTTGA 1425  
Db 1398 TGATTAATAAAGGTGACAGAGTGTATTTGAGGCTTTGAGATTAACCTGTGGAGACATCAA 1457  
Qy 1426 TGATGAATCGACTATATAAGAAAATTGTAGATGATTTATGTGCATTTCTAGCAACAAT 1485  
Db 1458 GGAATGTCCCAAGTGATTAAGTCAAGTATGAGATATTTCTGCTTCTAGCTCCGAT 1517  
Qy 1486 TAACCATCCAGAGGACTTGGCAAAACCAATTCCTCAATGAGATATCTGAAGCAGAGT 1545  
Db 1518 TCGTATCTCAGAACGCTTTAGCAAAACCAATTCGCAATTTACTACACTGATGTGAGAT 1577  
Qy 1546 TCGTATTTGCTCAATTTAGCTGATTAAGTATACAACTCAGATGTGTACATTTTGTAGAAC 1605  
Db 1578 TCAAGTAGCAAGTTGGCAGGCAAGTACACACAGAAAGAGGTTATATCTTGTATCTCTCG 1637  
Qy 1606 TGATATATACGTATGAAGAGATGCAATATGTATACGCTCATATGGGCCATATGACTG 1665  
Db 1638 TGATATACCAAGTATGAGGGGATGCTTATGTATCTCCATCTCAATGTACCCATACCCACTG 1697  
Qy 1666 GATTGGAAGATATGACCTTCTGATTAAGAAAAGTTGCAAGCTCAACCTATACTAAGA 1725  
Db 1698 GATTAAAAAAGATATGTTGTCTGAAGCTGAGAGCGGACGACCCAGCTTATGTGTAAGA 1757  
Qy 1726 AAAAGTATCTTACCTCATCTCCAGACGAGATGTTAAAGCAATCCAACTGGAGATAG 1785  
Db 1758 GAAAGTTTGACCCCTCTCTCGACAGACATCAGATTTCAAGAAATCTGAGCGCAAAAG 1817  
Qy 1786 TGCAGCAGCTATTTACATGCTGTGAAGAGGAAAACAAATTCACCTCGTTGCACTTC 1845  
Db 1818 AGCAGAAAGCTATCTACACCGCGGAAAGCAAGCTTAAAGGTGCACTGTATCTATGCG 1877  
Qy 1846 ATATATGTTGAGCATCAGTGTAGGTTAAAGAGTAAATTTGATTTCTCTCATTAAGA 1905  
Db 1878 TTACATCTTCAATATATCTGTAGAGTCAAAACAGTAGTATTAATCAATCTCATTTGA 1937  
Qy 1906 TCAATTACCATTAATTTAAATTTGCTTGTGTGATGATCACACATACAAAGCTCCAAATGG 1965  
Db 1938 CCATTTACCATTAATCAATTAATTTGAGTGTGTTGAGGAAGGCTTATATGAGCACTTAAGGG 1997  
Qy 1966 CTATACCTTGAAGATTTGTTTGGCAGATTAATGATCTAGTAAACACCTTGACGAGC 2025  
Db 1998 GTATACCTCTTGAAGATCTTTTGGGCACTGTCAAGTACTATGTGCAACATCCAAACGAGC 2057  
Qy 2026 TCCCATTTCAATGATGATGAGGCAATGCGCAGTGAAGCATGTGTAGSCAAAGAAAGCA 2085  
Db 2058 TCCGATTTCAAGATATATGTTTGTGTAACGCTAGCAGCATGTTCAAAAGAAACAAAATGG 2117  
Qy 2086 CAGTGAAGATCCAAATTAAGAACTTCAAGCGGATGA----- 2121  
Db 2118 TCAAGCTGATATCAATCAACGGAAGAAACCAAGAGAGAGAAACCTCAGACAGAAAACC 2177  
Qy 2122 ----- 2121  
Db 2178 TGAGAGAAAACCCCTGAGAGAGAAAACCGCAAGGAGAAACAGAGTCTCCAAAACC 2237  
Qy 2122 ---AGACCCAGTAGAGAAACACCTGCTGAGCCAGGAAGTCCCTCAAGTAGAGACTGAAA 2178  
Db 2238 AACAGAGAAACCCAGAAATTCACAGAGAAATCAGAAAGAACTTGAAGTCCGAGACTGAAA 2237  
Qy 2179 AGTAGAAGCCCAACTCAAGAGAGAGAAAGTGTGCTGCGAAGTAAAGCAATTTAGTCT 2238  
Db 2298 GGTGGAAGAAAAGTGAAGAGGCTGAAGATTTACTTGAAAAAATTCAGAGATCCAAATAT 2357  
Qy 2239 GAAAGCCAAATGCAACAGAACTTACGTGTTTACGAATATATTTGACTCTTCAATATAT 2298  
Db 2358 CAAGTCCAAATGCCAAAGAGACTCTCACAGATTTAAAAAATTAATTTACTATTTGGCACCA 2417



Db 1218 TCGTATATCCAGCCAGAGATCTTTCAGCAGAAACAGCAGCGATGTATGACAAACT 1217  
OY 1246 ATCAAAACAGAGAGTGTTCACACACTTACGTGCTAAAAAGAAATGTGCTCGC 1305  
Db 1278 GGCCAAAGCAGGAAGTTTATCTATAGCTAGAGACTAAGAAACAGACCTCCCATCTAG 1337  
OY 1306 TGACCAAGATTTTATGATTAAGCATTAATCTGTTAACTGAGGCTCATAAAGCCTGTT 1365  
Db 1338 TGATCGAGATTTTACAAATAAGGCTTATGACTACTAGCAAGAATATCACCAGATTTACT 1397  
OY 1366 TGAATTAAGGCTGCTAATCTGATTTCCAGCCTTAGAACAAATATTAAGCCTTGAA 1425  
Db 1398 TGATTAATAAGGCTGAGCAAGCTTATTTGAGCTTTGGATACCTGTTGAGACAGCAAA 1457  
OY 1426 TGATGAATGCACTAATTAAGAAAAATGTAGATGATTTTGGCATTCCTAGACCAAT 1485  
Db 1458 GGAATGCCAAGATGATTAAGCTCAAGTTAGTATGATTTCTTGCTTCTTAGCTCCGAT 1517  
OY 1486 TACCATCCAGAGCGACTTGGCAAAACCAATTTCTCAAAATGAGTATATCTAGACGAGT 1545  
Db 1518 TCGTCATCCAGAACGTTTAGAAACCAATGCGCAAAATTAACCTAGATGATGAT 1577  
OY 1546 TCGTATTTGCTCAATAGCTATAGTATAGACGCTAGATGTTTACTTTTGTATGACA 1605  
Db 1578 TCAAGTAGCCAAAGTGGCAGGCAAGTACACAGAAAGAGGTTATATCTTTGATCTCTG 1637  
OY 1606 TGATATATGATGATGAAGAGATGATGATGATGATGATGATGATGATGATGATGATG 1665  
Db 1638 TGATATTAACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1697  
OY 1666 GATTGGAAGATAGGCTTCTGATGAAGAAAAAGTTGCAAGCTCAAGCTTATCTAAAGA 1725  
Db 1698 GATTAAAAAGATAGTGTGCTGAGCTAGAGAGCGGCAAGCCAGCTTATGATAGA 1757  
OY 1726 AAAAGTATCTCTACCTCATCTCCAGACGAGATTTAAAGCAATCAACTGAGATAG 1785  
Db 1758 GAAAGTTTGACCCCTCTTCGACAGCATCGAGATTCAGAAATTAAGTGAAGCAAAAGG 1817  
OY 1786 TGCAGCAGCTATTTTACATAGTGTGAAGGGGAAAGAAAGCAATTCAGCTGCTGACTTC 1845  
Db 1818 AGCAGAAAGCTATCTACACGCGCTGAAGACGCTAAGAAAGTGCCTTATGCTATGCC 1877  
OY 1846 ATATATGTTGAGCATACAGTATGAGTTAAAAAGGTAATTTGATTAATCTCATAGA 1905  
Db 1878 TTACAAATCTCATATATCTAGTAAGTCAAAAGCGTATTAATCATCTATCTATGTA 1937  
OY 1906 TCAATTACATTAATTAATTTGCTGTTGTTGATGATACACATACAAAGCTCCAAATG 1965  
Db 1938 CCATTACCATTAACATCAAAATTTAGTGGTTTGACGAAGGCTTTATGAGGCACCTAAGG 1997  
OY 1966 CTTATACCTTGAAGATTTGTTTGGCAGATTAAGTACTAGTAAGACACCTGACGAG 2025  
Db 1998 GATTAATCTTGAAGATCTTTTGGCAGCTGCAAGTACTAATGTAACATCAAGCAAG 2057  
OY 2026 TCCCAATTTCAATGATGATGAGGCAATGCGCAATGAGCATGTTAGGCAAGAAAGCA 2085  
Db 2058 TCCGCAATTCAGATTAATGTTTGTGTAAGCTAGACGACATGTTCAAGAAACAAAAATG 2117  
OY 2086 CACTGAAGATCAATTAAGAACTTCAAGCGGATGA----- 2121  
Db 2118 TCAAGCTGATTAACCAATCAAGCAAAACCAAGCAGAGAGAAACCTCAGACAGAAAAAC 2177  
OY 2122 ----- 2121  
Db 2178 TGAAGAAAGAAACCCCTGAGAGAGAAACCGCAAGCGAGAAACAGAGTCTCCAAAAAC 2237  
OY 2122 ---AGAGCCAGTAGAGAAACACCTGCTGAGCAGAAAGTCCCTCAAGTAGAGACTGAAAA 2178  
Db 2238 AACAGAGGAACAGGAAGAAATCACAGAGGATCAGAAAGAACTCAGTGCAGACTGAAAA 2297  
OY 2179 AGTAGAAGCCCAACTCAAGAGACAGAGATTTTGTGCGAAGATTAAGGATTTCTACTCT 2238  
Db 2298 GGTGTAAGAAAGAACTGAGAGAGGCTGAAGATTTAGTTGAAAAAATCCAGATCCAAATTA 2257

OY 2239 GAAAGCCATGCAACAGAACTAGCTGTTTGAATTAATTTGACTCTCAATTAAT 2298  
Db 2258 CAACTCAATGCCCCAAGAGACTCTCACAGATTAATAAATTAATTTACTATTTGGCACCA 2417  
OY 2299 GGAATTAACATTAATATCATGTCAGAGAGCAAAATTAATTCGCTTGTAAAGGAACTAA 2358  
Db 2418 GGACAAACATTAATTAATGTCAGAGAACTGAAAAAATCTTGTGCTTATTAAGGAGACTAA 2477

RESULT 10  
AAA47605 standard; DNA; 2531 BP.  
ID AAA47605:  
AC AAA47605:  
XX  
XX 20-OCT-2000 (first entry)  
DT  
XX  
XX Recombinant variant of Sp36 gene (Sp36B) of *S. pneumoniae*.  
DE  
XX Streptococcus pneumoniae; infection; vaccine; coiled coil region;  
KW histidine triad residue; Sp36; antibody; otitis media;  
KW nasopharyngeal infection; bronchial infection; bronchitis; sepsis;  
KW meningitis; lobar pneumonia; ds.  
XX  
OS Streptococcus pneumoniae.  
XX  
FH Key location/Qualifiers  
FT CDS 1..2531  
FT /\*tag= a  
FT /product= sp36B polypeptide  
PN W0200037105-A2.  
XX  
XX 29-JUN-2000.  
PD  
XX 21-DEC-1999; 99MO-US30390.  
PF  
XX 21-DEC-1998; 98US-0113048.  
PR  
XX (MED1-) MEDIMUNE INC.  
PA  
XX Johnson LS, Koenig S, Adamou JE;  
PI WPI: 2000-452129/39.  
DR P-PSDB; AAB01469.  
DR  
XX  
XX Vaccine useful for prophylaxis and treatment of pneumococcal infections  
PT such as otitis media, nasopharyngeal and bronchial infections,  
PT comprises Streptococcus pneumoniae proteins  
XX  
XX  
PS Disclosure; Page 69-70; 70pp; English.  
XX  
XX Although a number of proteins have been suggested as being involved  
CC in the pathogenicity of Streptococcus pneumoniae, there still remains  
CC a need to identify polypeptides having epitopes in common from  
CC various strains of *S. pneumoniae* in order to utilize such  
CC polypeptides in vaccines to protect against a wide variety of  
CC *S. pneumoniae*. New vaccine compositions are described which comprise a  
CC Streptococcus pneumoniae polypeptide (or fragments) of 80 - 680 amino  
CC acids in length that comprise at least one histidine triad residue  
CC (HxxHxH) or a coiled-coil region, or an antibody directed against  
CC these features. The vaccine is useful in protecting against infection  
CC by Streptococcus pneumoniae. The vaccine composition comprising  
CC antibodies to is useful for passive immunization for treating  
CC Pneumococcal infections which includes otitis media, nasopharyngeal  
CC and bronchial infections.  
XX  
SO Sequence 2531 BP; 861 A; 508 C; 550 G; 609 T; 3 other;

Query Match 41.5%; Score 991; DB 21; Length 2531;  
Best Local Similarity 65.7%; Pred. No. 6,6e-239;  
Matches 1609; Conservative 1; Mismatches 697; Indels 141; Gaps 6;







Db 918 AACGGCAGAGGTGTAGCTGTCCCTCATGTGTAACATTTACCACTTATTCCTTATGAACA 977  
QY 958 AATGCTGAATTTGGAGAAGAAATCGCTGTATTTATCCCTTCGTTATGCTTCAACCA 1017  
Db 978 AATGCTGAATTTGGAAAAAGAAATGCTGTATTTATCCCTTCGTTATGCTTCAACCA 1037  
QY 1018 TTGGGTACAGATTTCAAGGCCAGAACACCAGTCCAAACGAGTCCGAGAACCTAGTCC 1077  
Db 1038 TTGGGTACAGATTTCAAGAACCAACCAAGTCCCAATTCAGATCCGAGAACCTAGTCC 1097  
QY 1078 AGGCCCCGAAACCTGCAACCAATCTTAAATATGACTGAATTCCTC-----TTT 1125  
Db 1098 AAGTCCGCAACCTGCAACCAATCTTCAACCAAGCAATTCATTTATGATGAGAAATTT 1157  
QY 1126 GGTATGATCAGCTGTAGCAAAAGTTGGGAGAGATATGTTTGAAGAAAAAGGCAATCTC 1185  
Db 1158 GGTCAAGAAAGCTGTGCAAAATAGCGATGTTATGCTTTGAGAGAGATGAGAGTTTC 1217  
QY 1186 TCGTTATGCTTTGCGAAGATTTACCATCTGAATCTGTTAAATCTTGAACCAATTT 1245  
Db 1218 TCGTTATGCTTTGCGAAGATTTACCATCTGAGCAAGACAGCAGCATTTGATGCAAACT 1277  
QY 1246 ATCAAAAAGAGAGTGTTCACACACTTAACTGCTAAAAAAGAAATTTGCTCTCTCG 1305  
Db 1278 GGGCAAGCAGAAAGTTTATCTATAGCTTAGAGCTAAGAAACTGACCTCCATCTAG 1337  
QY 1306 TGACCAAGAAATTTTATGATTAAGCATATATCTGTTAACTGAGAGCTCATTAAGCTTGT 1365  
Db 1338 TGATTCGAGAAATTTTACATATAGCTTATGACTTACTAGCAAGATTCACCAATTTACT 1397  
QY 1366 TGAATAATGAAGGTGCTAATTTCTGATTTCCAGCCTTAGACAAATTTATAGACGCTTGA 1425  
Db 1398 TGATTAATGAAGGTGCTAATTTCTGATTTTGAAGCTTTGATTAACCTGTTGGAACGACTCA 1457  
QY 1426 TGATGATGACTAATTAAGAAAAATTTGATGATGATTTTGGCAFTTCTTACACCAAT 1485  
Db 1458 GGGATGTCAGATGATTAAGTCAAGTTAGTGATGATTTCTTCTTACCTCCGAT 1517  
QY 1486 TACCCATCCAGAGCAGCTTGGCAACCAATTTCTCAATTAAGTATTAAGTAAGCAAGT 1545  
Db 1518 TCGTCATCCAGAGCTTTTGAAGAAACCAATTTGCGCAATTTACTACATGATGATGAT 1577  
QY 1546 TCGTATTTGCTCAATTTAGCTGATTAAGTATTAAGCAAGTGTATACATTTTGTATGAACA 1605  
Db 1578 TCAAGTAGCCCAAGTGGCAGGCAAGTACACAGAAAGGTTATATCTTATCTCTCG 1637  
QY 1606 TGATTAATGAGATGAAGAGAGATGATATGTAAGCCCTATATGGCCATGATGACTAG 1665  
Db 1638 TGATTAATGAGATGAAGAGAGATGATGTAAGCTTCCATATGACCCATGACCACTG 1697  
QY 1666 GATTGAAAAAGATAGCCTTTCTGATTAAGAAAAAGTTGCACTCAAGCCTTATCTAAGA 1725  
Db 1698 GATTAAAAAAGATAGTTTGTCTGAGCTGAGAGCGGAGCCGACGCTTATCTAAGA 1757  
QY 1726 AAAAGTATCTTACCTTCATCTCCAGACGAGATGTTAAAGCAATTCACACTGGAGATAG 1785  
Db 1758 GAAAGGTTTACCCCTCTTCGACAGACCATCAGATTTCAGAAATATCTGAGCAAAAG 1817  
QY 1786 TCGAGAGCATTTTCAATGCTGTGAAGGGAAAAAGCAATTCCTGCTGTCGACTTCC 1845  
Db 1818 AGGAGAGCATTTTCAACCGCGTGAAGAGAGCTGAAGAGCTGATGATGATGCTG 1877  
QY 1846 ATATATGAGTTGAGCATACAGTTGAGGTTAAAAAGCGTAAATTTGATTAATTCCTCATAGA 1905  
Db 1878 TTACATCTTTCAATATACCTGTAGAAAGTCAAAACGGATGTTATATCATATCATATATGA 1937  
QY 1906 TCATTAACATATATTAATTTGCTGTTGTTGATGATCACATATACAAAGCTCCAAATG 1965  
Db 1938 CCATTAACATATATTAATTTGCTGTTGTTGATGATGATGATGATGATGATGATGATG 1997  
QY 1966 CTATACCTTGAAGATTTGTTTGGCAGATTAAGTACTAGTGAACACCTTACGACAG 2025  
Db 1998 GATATCTTGTAGGATCTTTTGGCAGCTGCAAGTACTATGTGAAACATCCAAACGAAAG 2057

QY 2026 TCCACATTTCTATGATGATGAGGCAATGCAATGAGCATGTGTAGGCAAGAAAGACCA 2085  
Db 2058 TCCGATTTGAGATATGTTTGTGTAACCTAGCGACCATGTTCTGTAATAATTAAGGTAA 2117  
QY 2086 CAGTGAAGATTCATAATTAAGACTTCAAGCGGATGAGAGCCAGTAGAGGAAACACCTGC 2145  
Db 2118 CCAAGACAGTAAACCTGATGAAAGATGAAGTGAAGTAAAGTGAAGTGAAGTGAAGTGA 2177  
QY 2146 TGAGCAGAAAG 2156  
Db 2178 TGATCTGATG 2188

RESULT 12  
AAV27356  
ID AAV27356 standard; DNA; 2290 BP.  
XX  
AC AAV27356;  
XX  
DT 02-OCT-1998 (first entry)  
XX  
DE Streptococcus pneumoniae SP0042 nucleotide.  
XX  
KW Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;  
KW detection; pneumonia; otitis media; meningitis; ss.  
XX Streptococcus pneumoniae.  
XX  
FH key  
FT CDS  
FT 2..2290  
FT /tag= a  
FT /product= "SP0042"  
FT /transl\_except= (pos:152..154,aa:Xaa)  
FT /transl\_except= (pos:1406..1408,aa:Xaa)  
FT /transl\_except= (pos:1430..1432,aa:Xaa)  
FT /note= "no stop codon given; Xaa is unspecified"  
XX  
XX  
XX W09818930-A2.  
XX  
XX PD 07-MAY-1998.  
XX  
XX PF 30-OCT-1997; 97WO-US19422.  
XX  
XX PR 31-OCT-1996; 96US-0029960.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX PI Chol GH, Hromocky J A, Johnson LS, Kunsch CA;  
XX  
XX DR WPI; 1998-272224/24.  
XX  
XX P-PSDB; AAM55095.  
XX  
XX PT Nucleic acid encoding antigenic peptide(s) from Streptococcus  
XX pneumoniae - or their epitope-containing fragments, useful in  
XX protective or therapeutic vaccines, and for diagnosis  
XX  
XX PS Claim 1: Page 61-62; 118bp; English.  
XX  
XX CC The present sequence encodes a protein from Streptococcus pneumoniae.  
XX CC The nucleic acid sequence encoding the Streptococcus pneumoniae protein  
XX CC can be useful in vaccines for inducing protective antibodies against  
XX CC Streptococcus pneumoniae, for treatment or prevention of infection e.g.  
XX CC pneumonia, otitis media or meningitis. Probes based on the nucleic acid  
XX CC are used to detect Streptococcus infection (by usual hybridisation or  
XX CC amplification methods), also for isolating Streptococcus genes or their  
XX CC allelic variants. The protein can be used similarly to detect specific  
XX CC antibodies in standard immunoassays, especially for diagnosing or  
XX CC monitoring infections. Antibodies which bind the protein are used to  
XX CC detect corresponding antigens, to purify the protein and for passive  
XX CC immunisation (optionally coupled to a toxin). Vaccines are administered,  
XX CC e.g. by injection, orally or through the skin, typically at 0.01-1000  
XX CC (especially 10-300) mu g/ml per dose.

XX	Sequence
SQ	2290 BP; 766 A; 474 C; 498 G; 547 T; 5 other;

Query Match	41.3%	Score 987.6;	DB 19;	Length 2290;
Best Local Similarity	67.7%	Pred. No. 4.5e-238;		
Matches 1481; Conservative	0;	Mismatches 645;	Indels 60;	Gaps 5;

QY	1	TTTACGAGTGGGAGCTGTAATACGCTGAACGGTTAAGGAAAA---AAATCGCTTTC	57
Db	4	TTCTCTAAGAACTGGTGTGTGCACCAAGCTGGTCAGGTTTAAAGAAAGCTCTAACTCGAGTTTC	63
QY	58	CTATATAGATGAGAAAACAAACGACGAGCAAAAAGGAGAAATTGACTCCTGATGAGAGTAG	117
Db	64	TTATATAGATGATGATCAGGCTGGTCAAAAAGCAGAAAACCTTGACCCAGATGAAGTCAG	123
QY	118	CAACCGTGAAGGAATCAATCTGAGCAAACTCGTCATCAAGATTAACAGACCAAGCTATGT	177
Db	124	TAAAGGAGGAGGGGATCAACCCGCAACCAAAATNGTAACTAAGATTAAGCATCAAGGTTATGT	183
QY	178	CACATTCACATGCGCAGCACCATCATATTATATTCATAGGTAAGTTCCTTATGACGCTATCAT	237
Db	184	GACCTCTCATGAGACCATATTATCATTAATTAATGCAAGGTTCCCTTATGATGCATCAT	243
QY	238	CAGTGAAGAATTACTCATGAAGAATCCAAACTTAAGCTPAAAGATGAGATATTTGTTAA	297
Db	244	CAGTGAAGAGCTCCTCATGAAGAATCCGAATTATCAGTTGAAGGATTCAGACATCTGCA	303
QY	298	TGAGGTCAAGGGGTGATATGTTTTCAGAGGATGTAAGTGAATAATCTATGTTTACTTTAAGGA	357
Db	304	TGAATATCAAGGGTGGTATTATCTCATTAAGGTAAGGTAATACTATGTTTACTTTAAGGA	363
QY	358	TGCTGCCACCGCGGATTAACCTCCGTACAAAAGAGAAATCAATCGACAAAACAGAGCA	417
Db	364	TGCAGCTCATGCGGATTAATATTCGACAAAAGAGATTTAAACGTCAGAACAGAGGAAG	423
QY	418	TATGTCACATCTGTGAAGGTGGAACCTCCAGAAAACGATGGTGTCTTGGCTTGGCACGTTG	477
Db	424	CAGTCATTAATCAT-----AAGTCBAAGCAGATAATGCTGTTCGACGACCGAGAGC	474
QY	478	GCAAGGAGCGTATACTACAGATGATGGTATATCTTAAATGCTTCGATATCATGAGAGA	537
Db	475	CCAAGGACGTTATATCAACGAGATGATGGTATATCTTCATGATCATCGATATATATTGAGA	534
QY	538	TACTGTGATGCTTATATGCTTCCCTCATGAGATCATTAACATTTACATTTCTAAGATGA	597
Db	535	CACGGGTGATGCTTATATCGTTCTCTCAGCGCGACATTCACATTCATCTCAAAAGAA	594
QY	598	GTTTATTCAGTAGGAGATGGCTGCTGCAAGAGGCTTCATCTGGTCAGAGAAATCTGTG	657
Db	595	GTTATTCAGCTTAGCGAGTTACTGCTGTGAGAGGCT-----	629
QY	658	AAATTCAGAACCTATGCGCGACAAAATAGCATTAACACTTCACAGAACAACTGGTGACC	717
Db	630	-----ATTGGAATGAGGAAGCAGGAGATCTGCTCTTCCTCAAGTTAGTTAATATGC	681
QY	718	TTTCTGTAAAGCAATTCAGAGACTACAAATACTAACACAAACCAACAGCAGACTTAACAG	777
Db	682	AAATTCAGCTCAACCAAGATTTGTCAAGAACCAACAAATCTGACTGCACCAACTATCA	741
QY	778	TCAAGCAGTCAAGATATATACATATATATGTCCTTGAAGACGCTCTACAAATCCCTTT	837
Db	742	TCA---AAATTCAGGGGAAAAACATTTTCAGGCCCTTTTACTGATATGTATAGCTAAACCTT	798
QY	838	GAGTCAACGACATGTAGATCTGTATGGCCCTGTGCTTTGATCCAGACCAAAATCACAAAGTCG	897
Db	799	ATCAGACAGCCATGTGTGAATCTGATGGCCCTTATTTTCGACCCAGCGCAAAATCACAAAGTCG	858
QY	898	AACGAGCTAGAGGTTTCCATGTGCCACACGAGAGTCACTTACCACTTCATCCCTTACTGTCA	957
Db	859	AACGCCAGAGGTGTACTGTCCCTTATGTGTAAACCATTAACACTTATCCCTTAATGAACA	918
QY	958	AATGTGAAATTGGAAAGCAAAATCGCTGCTATTAATCCCTTGGTTATCGTTTCAAAACA	1017

Db	919	AATGTCGTAATTTGGAAAAACGAATTGCTGTAATTAATTCCTTCCTTGGTATGCTTAAACCA	978
Qy	1018	TTGGGTACCGAGATTCAAGGGCCAGACAGCAAGTCCCAACCGCACTCGGAAACCTAGTCC	1077
Db	979	TTGGGTTACAGATTCAAGACCAAGACACCAAGTCCACATGTAGCTCGGAAACCTAGTCC	1038
Qy	1078	AGGCCCCGACACTGACACCAATCTTTAAATAGACTCAATTTCTTC-----TTT	1125
Db	1039	AAGTCCGCAACCTCCACCAATCTCCTAACACGCTCCACCAATCCAAATTGATGAGAAAT	1098
Qy	1126	GGTATGTCAGCTGGTACGAAAGTGGGGAGAGATGTATGTCGAAGAAAGGCGATCTC	1165
Db	1099	GCTCAAGAAAGCTGTTGGAAGATGAGCGATGGTATGTGCTTTGAGGAAATGAGGTTTC	1158
Qy	1186	TGCTATATGCTCTTCGGAATAATTTACCATCTCAACTGTATAAAATCTTTGAAGCAAGT	1245
Db	1159	TGCTATATATCCAGCCCAAGATCTTTTCAGAGAAACAGACGAGCGATTTGATGAAACT	1218
Qy	1246	ATCAAAACAGAGAGTGTTCACACACTTTAACTGCTAAAAAAAGAAATGTTGCTCTCG	1305
Db	1219	GGCCACGAGGAAAGTTTATCTCATAGCTAGAGCTAGAAAGAACTGACCTCCCTATG	1278
Qy	1306	TGACACCAAGANTTTATATATAACCATATATCTGTTAACTAGAGCTCATAAAGCTTGT	1365
Db	1279	TGATGCGAAATTTTATTCATATAGGCTTATGCTTACGTAAGAGAAATTTACCAAGATTACT	1338
Qy	1366	TGNAATATPAAAGGTGCTGAATTCTGATTTCCAGCCTTAGACAAATTAATGAAAGCTGTAA	1425
Db	1339	TGATATATPAAAGGTGACAACTTGATTTTGAGGCTTTGGATTAACCTGTTGGAAACGCTAA	1388
Qy	1426	TGATGAATCGACTATPATAAGAAAAAATTTGGTAGATTTATTTGGCATTCAGACCAAT	1485
Db	1399	GGATGTGCAAGATGATTAAGTCAAGTTAGTGANGATATCTTGCTTGTAGTCTCGAT	1458
Qy	1486	TACCATTCAGAGGACTGTGGCAACCAATTTCTCAATTTGAGTACTGAAGACGAAT	1545
Db	1459	TGCTGATCGACAAAGGTTTAGAAAAACCAATGGCCAAATTTACCTCACTGATGATGAGAT	1518
Qy	1546	TCGATTTGCTCAATTACTCTGATPAGTATPACAAAGTCAGATGGTTACATTTTGTATGAGCA	1605
Db	1519	TCAGTATGCCAAGTTGGCAGGCAAGTACACAGAGAGCGGTTATATCTTTGATCTCTCG	1578
Qy	1606	TGATATATCAGTGTAGAGAGAGATGCATATGTAAACGCTCTATATGGCCATAGTCACTG	1665
Db	1579	TGATATATACAGTATATAGGGGGATGCGTATGTAACTCCACATATAGCOCATAGGCACCTG	1638
Qy	1666	GATTGAAAAAGATAGCCTTTCTGTATPAAAGAAAAAGTTGCAGCTCAAGCCTATATAAGA	1725
Db	1639	GATTAAAAAAGATAGTTTGTCTGACCTGAGAGCGGCGCCAGGCTTATGTCTTAAAGA	1698
Qy	1726	AAAAGGTAATCCTACCTCCACTCTCAGACCGCAAGTGTAAAGCAAAATCCAACTGGAGATAG	1785
Db	1699	GAAAGGTTTGACCCCTCTCTTCGACACACATCGACGATTTAGGAAAAATCTGAGGCAAAAGG	1758
Qy	1786	TGCAGCAGCTATTATTCATGTGTGAAGGGGAAAAAACAATTCACCTGTTGCACTCC	1845
Db	1759	AGCAGAAAGCTATCTACAAACGCGCTGAAGAGCTTAAGAAGGTGCCACTTGATCGTATGCC	1818
Qy	1846	ATATATGTTGAGCATACAGTTGAGGTTAAAAACGGTAATTTGATTAATTTCTCATPAAAGA	1905
Db	1819	TTACAAATCTTCATATATCTAGAAAGTCAAAAAACGGTAGTTTATATCATACCTCATTAAGA	1878
Qy	1906	TCATTTACCATATATTTAAATTTGCTTGTTGTATGATCTACATPACAAAGCTCCAAATGG	1965
Db	1879	CCATTTACCATATACCTCAAAATTTTGAAGTGTTCGCAAGGCGCTTTTATGAGGCACTPAAAGG	1938
Qy	1966	CTATACCTGGAGAGATTTGTTTGCAGAGATTAAGTCTACGTAGAGAACCCCTGAGAGAGC	2025
Db	1939	GTAATACCTTGAGAGATCTTTTGGCGACTGTCAAGTACTATGTGCAACATCCAAAGCAAGC	1998
Qy	2026	TCACATTTCTAATGATGATGGGGCAATGCCAGTGACATGTGTTAGGCAAGAAAGACCA	2085
Db	1999	TCCCATTTAGATATATGTTTGGTAAAGCTAGCGCACTGTTCAAAAGCAAAATATGG	2058



Db	1099	GGCTAAAGAAAGCTGTTGGAANAAGTACGGCATGGTTATGTGCTTTGGAGGAAATGGAGTTTC	1158
Oy	1186	TCGTTATGTCTTTGCGAAGATTTTACCATCTGAACCTGTTAAAAATCTTGAAGCAAGTT	1245
Db	1159	TCGTTATATCCCAAGCCCAAGAGATCTCTTACAGAGAACAACAGCAGACGGCATTTGATAGCAAACT	1218
Oy	1246	ATCAAAACAGAGAGTGTTCACACACTTTAACTGCTAAAAAAGAAAATGTTGCTCTCG	1305
Db	1219	GGCCACACAGGAAAAGTTTATCTCTATAGCTAGAGAGTAAAGAAAACCTGACCTCCCATCTAG	1278
Oy	1306	TGACCAAGAATTTTATGATATTAACCATATATATGCTTTAACTAGAGCTCAATTAAGCCTGT	1365
Db	1279	TGATCGGAATTTTACATATAGGCTTATGACTTACTAGACAGAAATTCACCAAGATTTACT	1338
Oy	1366	TGNNAAATPAAGGGTCTGTAAATCTGATTTTCCAGCCCTAGACAAATTTATAGAAGCCTTGA	1425
Db	1339	TGATATATTAAGGTGCGACAAAGTTATTTTGGAGCTTTGGATATACCTGTTGGAAACACTCA	1398
Oy	1426	TGATGAATCGACTATATAAAGAAAAATTTGGTAGATTTATTTGGCATTCTAGACCAAT	1485
Db	1399	GGATGTCTCNCAAGTGATTAAGTCAAGTTAGTGANGATTTTGGCCTCTTAGCCGAT	1458
Oy	1486	TACCATTCACAGAGGACTTTGGCAAACCAATTCATAATTTGAGTATCTAGAGAGAACT	1545
Db	1459	TCGTCATTCAGAAAGCTTTAGGAAAACCAATGCGCAAAATTACCTCACTGATGATGAGAT	1518
Oy	1546	TGCTATTGCTCAATTTAGCTATATAGTATCAAGTCAAGCTCAGATGGTTTCAATTTTGGATGAACA	1605
Db	1519	TCAAGTACGCAAGTTTGGCAGGCAAGTACACACAGACAGCGGTTATATCTTTGATCCTCG	1578
Oy	1606	TGATATTAATCAGTATGAAGAGAGATCATATGTAAACGCTCATATGAGGCCATAGTCACTG	1665
Db	1579	TGATATTAACCAAGTATAGAGGGAGTGCCTATGTAACCTCCATATGACCATTAGCCATAGCACTG	1638
Oy	1666	GATTGGAAAAAGTATAGCCTTTCTATATAGGAAAAAGTTGCACTCAAGCTTATCTAAAGA	1725
Db	1639	GATTAAAAAAGATAGTTTGTCTGAACCTAGAGAGGCGGACGCCAGGCTTATGCTAAAGA	1698
Oy	1726	AAAAGTATCCTCAACCTCCATCTCCAGACCGACAGTGTAAAGCAATTCACACTGGAGATAG	1785
Db	1699	GAAGGTTTGAACCCCTCTCTTCGACACACATCGATCGATTCAGGAAATTAAGGCAAAAGG	1758
Oy	1786	TGCAGACGCTATTTTACATCGTGTGAAGGGGAAAAACGAATTCACCTCGTTGCACTTCC	1845
Db	1759	AGCAGAAAGCTATCTACAAACCGCGGTGAAGACGCTAAGAAAGTGCACCTTGATCCGATGCC	1818
Oy	1846	ATATATGCTTGAGCATACAGTTGAGGTTAAAAACGGTAATTTGATTTATTTCTCATPAAGA	1905
Db	1819	TTACAAATCTTCATATATCTACTGTAGAAAGTCAAAAAACGGTAACTTTAATCATCTACTTATAGA	1878
Oy	1906	TCATTACCATATATTAATTAATTTTCTGTTGTATGATCTACATCAACATCAAGCTCCAATGG	1965
Db	1879	CCATTACCATATACCTCAAAATTTTAGTGTGTTTGAAGAAAGGCCCTTTTATGAGGACCTTAAGGG	1938
Oy	1966	CTATACCTTGAAGATTTGTTTTCGACAGATTAAGTATCTACGTAGAACAACCTGACGAACG	2025
Db	1939	GTAATACCTTGAGAGATCTTTTGGCAGACTGTCAAGTATCTATGTGCAACATCCAAACGAACG	1998
Oy	2026	TCACATTTCTAATGATGATGGGCGCAATGCCAGTGAACATGTGTTAGGCAAGAAAGACA	2085
Db	1999	TCCCATTTCAAGATTAATGTTTGTGTAACGCTAGCGACCATGTTTCAAGAAACAAAATATGG	2058
Oy	2086	CAGTGAAGATCCAAATTAAGACTTCAAACCGGATGAAGAGCCAGTAGAGGAAACACTGCG	2145
Db	2059	TCAGCTGATATCCAAATTCAAACGGGAAACCAAGCGAGGAACCTCAGACGAAAAAAC	2118
Oy	2146	TGAGCCAGAAGTCCCTCAAGTAGAGA 2171	
Db	2119	TGAGGAAGAAACCCCTCGAGAAAGAGA 2144	

ID	AAA65737
AC	AAA65737 standard; DNA; 2639 BP.
XX	
XX	AAA65737;
XX	
DT	21-NOV-2000 (first entry)
XX	
DE	Streptococcus pneumoniae BVH-11-2 gene SEQ ID NO:13.
XX	
KW	Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine; prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia; otitis media; pneumonia; immunisation; bactericidal; ds.
XX	
OS	Streptococcus pneumoniae.
XX	
PN	WO200039299-AZ.
XX	
PD	06-JUL-2000.
XX	
PF	20-DEC-1999; 99WO-CA01218.
XX	
PR	23-DEC-1998; 98US-0113800.
XX	
PA	(BIOC-) BIOCHEM PHARMA INC.
PI	Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
DR	WPI: 2000-452397/39.
P	P-PsDB; AAB12720.
PT	Streptococcal antigens useful for vaccinating against e.g. meningitis,
PS	otitis media, bacteremia and/or pneumonia -
XX	Example 6; Fig 16; 106pp; English.
CC	The present invention describes nucleic acids (I) encoding protein antigens (II) from Streptococcus pneumoniae. The protein antigens have bactericidal activity. The nucleic acids, encoding the protein antigens, may be used for the recombinant production of the proteins they encode. The protein antigens may then be used as vaccines for the prevention and treatment of Streptococcal infections in mammals (especially humans) which result in, e.g. meningitis, otitis media, bacteremia and/or pneumonia. The present sequence encodes the <i>S. pneumoniae</i> BVH-11-2 protein antigen.
SC	Sequence 2639 BP; 889 A; 518 C; 567 G; 665 T; 0 other;
	Query Match            41.1%; Score 980.8; DB 21; Length 2639; Best Local Similarity 67.8%; Pred. No. 2.5e-236; Matches 1473; Conservative 0; Mismatches 638; Indels 60; Gaps 5
OY	1 TTCTTACAAGTTGGCAGCTATTCAAAGCTAAGAAGGTTAAGGAATA--TAATCGTGTTTC 57
DB	173 TTCATTAGACTGTGCCTCACCAAGCTGGTCAGGTTAAGAAAGTCTAAATCGAGTTTC 232
OY	58 CATPAPAGATGAAAAACAAGCAGCAGCAAACGGAATTGACGCCGATGAGGTAG 117
DB	233 TTAATTAGATTGTGATFCACGGCTGCTGTTAAAGCAGAAATTTTGACCACGATGAAGTCA 292
OY	118 CAAGCGTGAGAAATCAATCATGTGAGCAATTCATCAAGATACAGAACCAAGCGTATGT 177
DB	293 TAAGAGAGAGGGGATCAACGCCGACAAATTTTATCAAGATTTACGAGATCAAGTTATGT 352
OY	178 CACTTCACATGGCGACCACTATCATTTATTAACAATGGTAAAGTTCTTATGACGATCAT 237
DB	353 GACCTCTCATGAGACCATATFATCTACTATATGCAAGTTCCTTATGATGCATCAT 412
OY	238 CAGTGAAGATTAACATGAAGAATCCAACATATAAGCTAAAGAAAGATGAGATATGTTAA 297
DB	413 CAGTGAAGACTTCATCATGAAGAATCCGATATATCATGTTGAAGANTCAGACATTTGTCAA 472
OY	298 TGAGTCAAGGCTGATATGTTATCAAGGTAGATGAATAATCATATGTTACCTTAAGA 357

Db 473 TGAATCAAGGGTGGCTATGTGATTAAGTAGACGAAAAAATCTATGTACCTTTAAAGA 532  
QY 358 TGGTCCCGCCAGGATTAACGTCCTGACAAAAAGAGAAATCAATGCAGAAAAACAGAGCA 417  
Db 533 TGGGGCCCATGCGGACAAATTTGCGACAAAAAGAGATTTAAAGCTGAGAGAGAGACA 592  
QY 418 TAGTCAACATCGTGAAAGGTGGAACCTCCAGAAACGATGGTGGTGGTGGCAGCTTC 477  
Db 593 CAGTCATAATCAT-----AACTCAAGAGCATTAATGCTGCTGCAGCCAGCAGC 643  
QY 478 GCAAGGCGCTATACCTACAGATGCTGTTATGCTTAACTGCTTCGATATCATAGAGA 537  
Db 644 CCAAGGCGCTATACAGAGATGAGTGGTATATCTTCAATGCACTGTATCATGAGGA 703  
QY 538 TACTGTGATGCTTATATCTGCTCCATGGAGATCATTTACCATTTACCTTAAGATGA 597  
Db 704 CAGGGGTGATGCTTATATGCTTCTCCTCAGGCGACCATTTACATTTACCTTAAGATGA 763  
QY 598 GTTATCAGTACGAGAGTGGCTGCTGAGAAAGCTTCTATCTGGTGCAGGAAATCTGTC 657  
Db 764 GTTATCAGTACGAGAGTGGCTGCTGAGAAAGCT----- 798  
QY 658 AAATTCAGAACTATGCGCGAGCAAAATAGCGATTAACACTTCAAGAACAACTGGTACC 717  
Db 799 -----ATTGAAATGGAGAGAGGAGTCTGCTCTTCAAGTTCTAGTTATTAATGC 850  
QY 718 TTCTGTAGCAATCCAGAGACTACAAATACTAACACAGAACAGCAACACTAAG 777  
Db 851 AAATTCAGTTCAACAGATTTGTCAGAGAACCAATCTGACTGCTCCTCCAACTTATCA 910  
QY 778 TCAAGCAGTCAAGATTAATGATGATGCTCTTTGAAACAGCTTACAACTGCTTT 837  
Db 911 TCA---AAATCAAGGGGAAAAAATTCACAGCTTTTACGTAATGTTATCTTAAACCTT 967  
QY 838 GAGTCAACAGTATAGATCTGATGGCTTGTCTTGTATGATCGACAAATTCACAACTGC 897  
Db 968 ATAGAAAGCGCATGTAGAAATCTGATGGCTTATTTTGCACCGACGCAATTCACAACTGC 1027  
QY 898 AACAGCTAGAGGTGTGCACTGCGACACGAGATCATTTACACTTCACTTCCCTACTCTCA 957  
Db 1028 AACCGCAGAGGTGTGAGCTGCTCCCTCATGTTAACCATTTACACTTATGCCCTTATGACA 1087  
QY 958 AATGCTGAAATTTGGAAGAGCAATTCGCTGATTTATTCCTCTGTTATGTTCAACCA 1017  
Db 1088 AATGCTGAAATTTGGAAGAGCAATTCGCTGATTTATTCCTCTGTTATGTTCAACCA 1147  
QY 1018 TTGGGTACAGATTTCAAGGCGAGAACACAGCAAGTCCAGACGCTCCGAACTGATCC 1077  
Db 1148 TTGGGTACAGATTTCAAGGCGAGAACACAGCAAGTCCAGACGCTCCGAACTGATCC 1207  
QY 1078 AGGCCGCGCAACTGCAACCAATCTTAAATAGACTCAAAATTTCTTC-----TTT 1125  
Db 1208 AAGTCTGCAACCTGCAACCAATCTTCAACAGCTCCAGCAATTCGAACTCCGAACTGATCC 1267  
QY 1126 GGTATGCTGAGTGTGAGAAAGTTGGGAGAGATATGTTGAAAGAAAGGGCATCTC 1185  
Db 1268 GGTCAAGAAAGCTGTTGAAAGATAGCGAGTGTATGCTTTGAGAGAGATGAGATTTC 1327  
QY 1186 TCGTTATGCTTTGGCAAGATTTACCATGTAACCTGTTAAACCTTCAAGCAAGT 1245  
Db 1328 TCGTTATGCTTTGGCAAGATTTTTCAGAGAACAGCAGCGGATGATGCAAACT 1387  
QY 1246 ATCAAAACAAGAGATGTTTCAACACTTAACTGCTAAAAAGAAATGTTCTCTCTCG 1305  
Db 1388 GCGCAAGCAAGAAAGTTATCTATAGCTAGAGCTAAGAAAACTGACTCCCATCTAG 1447  
QY 1306 TGACCAAGAAATTTATGATTAAGCATATATCTGTTAATGAGGCTCATTAAGCCCTGTT 1365  
Db 1448 TGATCGAATTTTACAAATTAAGGCTTATGACTTACGAAAGATTCACCAAGATTTACT 1507  
QY 1366 TGAATAATAGGCTGCTATTTCTGATTTCCAGGCTTAGACAAATTTATAGAACGCTTGA 1425  
Db 1508 TGATTAATAAGGCTGCAAGGTTGATTTTGAGTTTGATTAACCTGTGTGAACGACTCAA 1567

QY 1426 TGATGAATGCACTAATTAAGAAAAATTTGTAGATGATTTATTTGGCATTCCTTACCAAT 1485  
Db 1568 GATGTCCTCAAGTATTAATTCAGATGATGATGATGATGATGATGATGATGATGATGAT 1627  
QY 1486 TACCATCCAGAGCACTTGGCAACCAATTTCTCAATTTGAATATGTAAGAGCAAGT 1545  
Db 1628 TCGTCACTCAGAAAGTTTAAAGAAACCAATTCGCAAAATTTACCTACATGATGATGAT 1687  
QY 1546 TCGTATGCTCAATTTAGCTGATAGTATACAGCTGATGATGATGATGATGATGATGAT 1605  
Db 1688 TCAAGTACAGAGTGGCAGGCAAGTACACAGAAAGAGGTTATATCTTATCTTATCCTCG 1747  
QY 1606 TGATATATCAGTATGAGAAAGAGATGATATTAAGGCTCATATGAGGCTATGACTG 1665  
Db 1748 TGATATATCAGTATGAGAGGAGATGCTATGATATCCATATGATGATGATGATGATGAT 1807  
QY 1666 GATTGCAAAAGATGACCTTTCTGATTAAGAAAAAGTTGACGCTCAAGCTTATCTAAGA 1725  
Db 1808 GATTAAAAAAGATGATGTTGCTGAGCTGAGAGAGGCGCAGGCTTATCTTAAAG 1867  
QY 1726 AAAAGTATCTACCTCCATCTCCAGCGAGATGTTAAAGCAAAATCCAACTGAGATAG 1785  
Db 1868 GAAAGGTTGACCCCTCTCTGACAGACACAGATTCGAAATTCGAGGCAAAAG 1927  
QY 1786 TGCAGCAGCTATTTACAAATGCTGTAAGGGAAGAAAAAGCAATTCACCTGCTGACTTC 1845  
Db 1928 AGCAGAAAGCTATCTACACCGCGTGAAGAGCTAAGAAAGTCCACTTATGATGATGCT 1987  
QY 1846 ATATATGTTGAGCATACAGTATGAGTTAAAAACGGTAATTTGATTTATCTCATAGGA 1905  
Db 1988 TTCAAAATCTTCAATATACGTATGAAAGTCAAAAACGGTATGTTATATCATCTTATATGA 2047  
QY 1906 TCATATCCATATATTAATTTGCTTGTGATGATCAACATATACAAAGCTCCAAATG 1965  
Db 2048 CCATTTACATTAATCAATTAATTTGATGATGATGATGATGATGATGATGATGATGAT 2107  
QY 1966 CTATACCTTGAAGATTTGTTGCGAGATTAAGTACTACGTAGAAACACCTGACGAG 2025  
Db 2108 GTATAGCTTGAAGATCTTTTGGCAGCTGCAAGTACTATGATGATGATGATGATGATGAT 2167  
QY 2026 TCCACATTTCAATGATGATGAGGCAATGCGACAGTACGATGATGATGATGATGATGAT 2085  
Db 2168 TCCGATTTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2227  
QY 2086 CAGTGAAGATCCAAATTAAGAACTTCAAGGCGATGAAGAGCCAGTGAAGAAACACTGC 2145  
Db 2228 CCAAGATGATTAACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2287  
QY 2146 TGAAGCCAGAAG 2156  
Db 2288 TGAATCTGATG 2298

RESULT 15  
ABK15104  
ID ABK15104 standard; DNA; 2639 BP.  
XX  
AC ABK15104;  
XX  
DT 08-MAY-2002 (first entry)  
XX  
DE DNA encoding Streptococcus pneumoniae BVH-11-2.  
XX  
KW BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteremia;  
KW pneumonia; streptococcal bacterial infection; gene; ds; BVH-11-2.  
OS Streptococcus pneumoniae.  
XX  
FH Key location/Qualifiers  
FT CDS 114..2630  
FT /tag= a  
FT /product= "BVH-11-2"



Db 1508 TGAATAAAGGTCGACAGTGTGAGGTTTGATTAACCTGTTGGAAAGACATCAA 1567  
QY 1426 TGATGATGCACTAATTAAGAAAAATGAGATGATTTATGSCATTCCTAGCACAAT 1485  
Db 1568 GGATGCTCAAGTGAATAGTGAAGTGAATGATGATGATGATGATGATGATGATGAT 1627  
QY 1486 TACCATCCAGAGGAGCTTGGCAACCAATTCCTCAATGAGTATGATGATGATGATGAT 1545  
Db 1628 TCGTCAATCAGAAAGCTTGAAGAAACCAATGCCAATTAACCTAGTATGATGATGAT 1687  
QY 1546 TCGATGCTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1605  
Db 1688 TCAAGTACCAAGTGGCAGGAGTACACAGAAAGAGAGGATTAATCTTGTGATCCTCG 1747  
QY 1606 TGATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1665  
Db 1748 TGATATATCAAGTGAATGAGGAGTGGCTATGATGATGATGATGATGATGATGATGAT 1807  
QY 1666 GATTGAAAGATAGCTCTTCTGATAAGAAAAAGTGCAGCTCAAGCTATATAAGA 1725  
Db 1808 GATTAAAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1867  
QY 1726 AAAAGTATCCTACCTCCATCTCCAGACGAGATGTTAAAGCAAAATCCAACTGAGATAG 1785  
Db 1868 GAAAGGTTTGACCCCTCTCCAGACGACCAAGATTCAGAAATATCTGAGGCAAAAGG 1927  
QY 1786 TGCAGCAGCTATTTACATAGTGTGAAGGGGAAAAAGAAATTCACCTGCTGACTTCC 1845  
Db 1928 AGCAGAGCTATCTACACGCGGAAAGCAAGCTAAAGAGTGCACCTGATGATGCC 1987  
QY 1846 ATATATGTTGAGCATACAGTGTGAGTTAAAGCGTAATTTGATATTCCTCATAGA 1905  
Db 1988 TTACAAATCTCAATATATCTAGAACGCAAAAAGGATGTTAATCATACCTCATATGA 2047  
QY 1906 TCATTTACCAATATTAATTTGCTGTTGATGATCACAATACAAGCTCCAAATGG 1965  
Db 2048 CCATTTACCAATACAATTTGAGTGTGAGCAAGGCTTTATGAGGACCTTAAGGG 2107  
QY 1966 CTATACCTTGGAGATTTGTTGGAGCATTAAGTACTAGTAGAACACCTGACGAAG 2025  
Db 2108 GTATAGCTTGAAGATTTTGGGAGCTGCAAGTACTATGTCGACATCCAAACGAAAG 2167  
QY 2026 TCCACATTTCAATGATGATGAGGCAATGCCAGTGAACATGTGTTAGCAAGAAAGACA 2085  
Db 2168 TCCGATTTCAAGTAATGTTTGTGTAACGCTAGTACCATGTGCTAAATAAGGCAGA 2227  
QY 2086 CAGTGAAGATCCAAATAGAACTTCAAGCGGATGAAGAGCCAGTAGAGAAACACCTGC 2145  
Db 2228 CCAAGATAGTAACTGATGAAGTAAGGAACATGATGAAGTAAAGTGAAGCCAACTCACCC 2287  
QY 2146 TGAGCCGAGAG 2156  
Db 2288 TGAATCTGATG 2298

Search completed: November 17, 2002, 01:50:41  
Job time : 373 secs